Contact Management Model Assessment: An Approach and System Description

S. C. Nardone
University of Massachusetts Dartmouth

D. J. Ferkinhoff
S. E. Hammel
K. F. Gong
Combat Systems Department

Naval Undersea Warfare Center Division
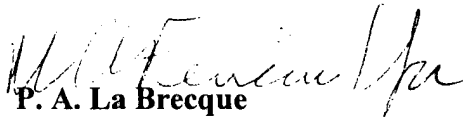# Newport, Rhode Island

19960313 104

PREFACE

This report was prepared under Project No. 33CWF55, "Submarine Combat Control Technology Task," principal investigator A. H. Silva (Code 2211). The sponsoring activity is the Office of Naval Research, program manager J. Fein (ONR-333).

The technical reviewer for this report was M. L. Graham (Code 2214).

The authors of this report gratefully acknowledge the support, assistance, and input of other personnel from Code 2211.

Reviewed and Approved: 10 April 1995

P. A. La Brecque
Head, Combat Systems Department

REPORT DOCUMENTATION PAGE

Public reporting for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 10 April 1995	3. REPORT TYPE AND DATES COVERED Final

4. TITLE AND SUBTITLE

Contact Management Model Assessment:
An Approach and System Description

5. FUNDING NUMBERS

6. AUTHOR(S)
S. C. Nardone
D. J. Ferkinhoff
S. E. Hammel
K. F. Gong

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Undersea Warfare Center Division
1176 Howell Street
Newport, Rhode Island 02841-1708

8. PERFORMING ORGANIZATION REPORT NUMBER

TR 10,444

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Contact Management Model Assessment (CMMA) is a problem of state estimation in the presence of modeling ambiguity and uncertainty. In the general context of CMMA, noise corrupted measurements from a contact are processed to obtain estimates of the contact's position and velocity. The process by which the contact state is estimated typically requires a model that relates contact dynamics, the sensor, and the environment to the measurements. In some instances, the tactical scenario may be such that these models are not uniquely known. Such conditions may exist in the littoral regions with multiple platforms operating in concert with one another. These situations give rise to conditions in which the actual process, that is its virtual analogy known as the model, is not known and multiple modeling hypotheses are possible. Classical hypothesis testing methodologies are applicable to CMMA but have the adverse effect of requiring multiple state estimates; some of which may experience convergence difficulties. The approach taken here uses Dempster-Shafer (DS) evidential reasoning to pre-select, for further processing, only those model hypotheses that are consistent with the observed data. In the DS approach, evidential reasoning takes place in the frame of discernment; a set of mutually exclusive and exhaustive hypotheses. Evidence, leading to belief in the subsets of the frame of discernment, is mapped into basic probability assignment over the power set of the frame. The evidence, or belief, from multiple sources that are related to the same frame of discernment can be combined with Dempster's combination rule. Dissimilar but related frames can be mapped to a common frame of discernment with compatibility relations. The collection of frames and compatibility relations is referred to as a gallery. This report details the CMMA gallery and provides a system description. An overview of DS evidential reasoning is presented along with the specific CMMA application. Functions that map evidence from statistical inferences on predicted residual sequences to belief in model hypotheses are developed. Additionally, a method for hypotheses selection from the DS frame is developed. Experimental results from computer simulations are presented and discussed.

14. SUBJECT TERMS			15. NUMBER OF PAGES 68
Underwater Tracking Target Motion Analysis	Evidential Reasoning Systems Detection Theory	Statistics	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CONTACT MANAGEMENT MODEL ASSESSMENT: AN APPROACH AND SYSTEM DESCRIPTION

1. INTRODUCTION

Tracking acoustic sources in the ocean environment remains a difficult problem and continues to be an area of active research.[1-4] In the contact management tracking problem, noise corrupted sonar measurements are used to obtain estimates of the contact's state parameters consisting of range, bearing, depth, course, and speed. The fundamental physical quantities measured by a sonar system are time delay and frequency. The time delay from spatially diverse sensor elements is often converted to angle of arrival information, examples of which are azimuthal bearing or conical angle depending on the sensor employed. State estimation in nonlinear dynamic systems of this type requires the data to be processed under various kine-matic, environmental and sensor modeling assumptions.[5-7]
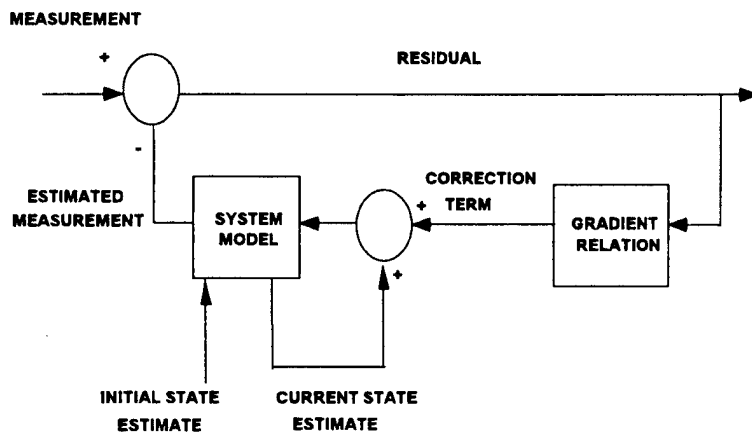
Figure 1. State Estimation Process

A typical state estimation process is depicted schematically in figure 1. As shown, the residual error in the fit to the measurements, or simply the residuals, is the difference between the actual measurements and an estimate of the measure-ments obtained from the state estimate and the system model. Using the residuals, the gradient relation maps the error in the fit to the data into a correction term for the state estimate. The gradient relation is based on the mathematical system model that relates the state to the measurements. The state estimate is updated with the correction term, new residuals are computed and the process is iterated until the estimation algorithm converges to a state estimate that provides a "best-fit," by some measure, to the observed measurements. Thus, because the estimated measurements and gradient relations are dependent on the model of the system, the judicious selection of appropriate models is critical if accurate state estimates are to be obtained. When the system model sufficiently reflects the actual process by which the measurements were produced, the residuals are noise-like in character. That is, they are devoid of any deterministic features.

Often the correct models are not known *a priori* and must be determined along with the contact state. Additionally, the system must also adapt to changing conditions of the scenario, that is, changes in the system model. Traditionally, these conditions are accommodated using a combination of batch or sequential state estimation techniques with multiple hypothesis test procedures.[8] In these approaches, all possible models are used to obtain state estimates and the

1

solution(s) with the best residuals or largest likelihood are selected as the correct model(s). As these model assessment methods apply all to models simultaneously, large amounts of computational resources may be required to achieve real time performance. Further, because some of the possible models are inconsistent with the observed data, the estimation algorithms may fail to converge. As a consequence, nontraditional methods for state estimation in the presence of model ambiguity and uncertainty are being explored for application to the contact management model assessment (CMMA) problem.

It is well known that changes in the underlying process can induce identifiable features into predicted residual sequences. A nontraditional processor for model assessment, which took advantage of this property, was originally proposed in Baylog, et al.[5] and extended in Ferkinhoff, et al.[9,10] The system, which continues to be developed, is illustrated in its currently implemented form in figure 2. The approach uses state estimation and hypothesis testing, coupled with Dempster-Shafer (DS) evidential reasoning,[11,12] to

Figure 2. Block Diagram of CMMA System

resolve modeling ambiguity. As different classes of modeling errors can induce unique features in residual sequences of the various measurement types, it is often possible in theory to deduce the type of modeling error, or at least identify the class to which it belongs. Previous reports have discussed the extraction and representation of evidence as well as the overall performance of the DS theory approach to the CMMA problem.[9,10] In those studies, it was assumed that modeling consistency had been achieved and that a change of the model subsequently occurred. However, the system is not limited to that application.

This report discusses several advances in the application of the DS theory of evidential reasoning to the CMMA problem. A hypothesis selection criterion for the DS frame of reasoning is developed. The approach uses a geometric interpretation of the DS frame ,which in itself can be a useful tool. With the development of the hypothesis selection criterion, the previous results are extended to include the efficient selection of models for further processing. Novel evidence representation functions that map evidence produced by probabilistic multiple hypothesis tests into belief in the deterministic features are developed. A two-stage approach is employed that first establishes the belief in the existence of a feature and then refines that belief to one of feature amplitude. This extension of the evidence representation allows for a greater degree of discernment which leads to higher resolution of the possible modeling hypotheses. These advances have been reported separately elsewhere[9,10,13,14] but are collectively presented here with the CMMA application for the first time.

2

The remainder of the report is organized as follows: section 2 reviews the DS theory of evidential reasoning so that the application to CMMA can be presented in the proper context. Included in this section are the geometric representation of the DS frame and the hypothesis selection criterion; section 3 details the application of the DS theory to the CMMA problem. Also included in this section are the general overview of the structure of the evidential reasoning process and the development of the evidence representations. Experimental results, obtained with simulated data, are presented and discussed for the CMMA problem in section 4 and section 5 contains conclusions with suggestions for future work. Several appendices are included that provide details of specific developments.

2. DS EVIDENTIAL REASONING

The DS theory of evidential reasoning is one of several approaches for producing inferences from uncertain information. Its appeal for application to model assessment is that it intrinsically accommodates the expression of ignorance and naturally provides a convenient framework on which the CMMA problem can be structured. Section 2.1 reviews the basic mathematical theory of DS evidential reasoning (DS-ER). A geometric representation, which is not a standard part of DS-ER, is presented in section 2.2. This representation is used in section 2.3 in the development of a hypothesis selection criterion.

2.1 DS THEORY

The basic structure for DS-ER is the frame of discernment. Denoted by Θ, the frame is a set of mutually exclusive and exhaustive hypotheses:

$$\Theta_A = \{a_1 ... a_N\} . \tag{1}$$

The power set ,denoted by 2^Θ, is the set of all subsets of the frame of discernment and has

$$K = 2^{|\Theta|} - 1 \tag{2}$$

elements, where $|\Theta|$ is the cardinality of the set Θ. The minus one accounts for the null set $\varnothing$ which is not considered a member of the power set in DS theory. Note that the cardinality of the power set grows exponentially with the number of hypotheses. Figure 3 illustrates a generic power set.
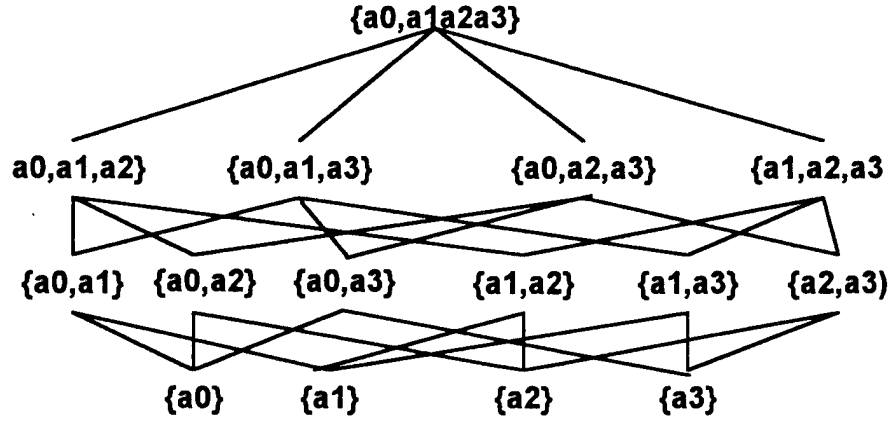
Figure 3. Power Set

A basic probability assignment (BPA), denoted $m(A_j)$, is a function that assigns a probability to each element of the power set, A_j, representing the belief that directly supports the truth of the hypotheses contained in the set A_j. The individual BPA's are bounded between zero and one and sum to unity. It is noted here that BPA has a special connotation in that it is defined to be beliefs assigned to the initial power set. While belief, which is generated in later stages of the system, is produced by operating on the original BPA, and is also referred to as $m(A_j)$, this belief is not generally referred to as BPA, only as belief. However, the belief distributed among the power sets in the latter sages of the system must still conform to the constraints imposed on the original BPA, i.e., they are bounded between zero and one and sum to unity over any power set. Additional measures of plausibility and support for each element can also be computed. The support for A_i, denoted by $s(A_i)$, is given by the sum of the belief over all the subsets of A_i as

$$s(A_i) = \sum_{a \subseteq A_i} m(a), \tag{3}$$

and represents the belief that directly supports the hypotheses contained in A_i. The plausibility of A_i, $p(A_i)$, is a measure of the lack of support in the complement of A_i and is given by one minus the support in the complement of A_i as

$$p(A_i) = 1.0 - s(\overline{A}_i) = \sum_{a \cap A_i \neq \varnothing} m(a). \tag{4}$$

Note that in general, support is less than or equal to plausibility and both are bounded by zero and one. The uncertainty in assigning belief to the hypotheses contained in element A_i is the difference between the plausibility and support. Note that DS theory includes Bayesian probability as a special case when all belief is distributed among the singleton sets $\{a_1\}, \{a_2\}, ..., \{a_N\}$.[11,12] Examples of computing plausibility and support are given at the end of this section and section 2.3.

4

If several bodies of evidence exist for a common frame of discernment, the resulting beliefs can be combined using Dempster's combination rule,

$$m_c(A_i) = \frac{1}{\alpha} \sum_{A_j \cap A_k = A_i} m_1(A_j)m_2(A_k), \tag{5}$$

where $m_1(A_j)$ and $m_2(A_k)$ are the belief in A_j and A_k generated from two bodies of evidence, E_1 and E_2, represented in two common frames of discernment Θ_1 and Θ_2, respectively. The term α is the renormalization constant necessary to account for belief being placed into the null intersection $\varnothing$ and is

$$\alpha = 1 - \sum_{A_j \cap A_k = \phi} m_1(A_j)m_2(A_k). \tag{6}$$

In both expressions, the indices j and k range over all the elements of the power set. Dempster's combination rule is both commutative and associative. An example using Dempster's combination rule is given at the end of this section.

Dempster's combination rule is valid only for identical frames of discernment. However, it is often the case that multiple bodies of evidence exist for dissimilar but related frames. To combine evidence from different frames of discernment it is necessary to map them to a common frame. This mapping is called a compatibility map and is a logical relationship (i.e., one or zero) between the hypotheses of one frame (A) and the hypotheses of another frame (B). Thus, compatibility maps, which are used to map belief in one frame of discernment to a different frame, are useful for combining belief that originates in dissimilar frames into belief in a common frame.[12,15] The compatibility map defines the relationship between the elements of two frames $\Theta_{A,B}$ where

$$\Theta_{A,B} \subseteq \Theta_A \times \Theta_B, \tag{7}$$

and the compatibility mapping is

$$C_{A \to B}(A_k) = \{b_j | (a_i, b_j) \in \Theta_{A,B}, a_i \in A_k\}, \tag{8}$$

where at least one pair (a_j, b_j) is specified for each of the A_k in Θ_A. The BPA for an element of the power set of the new frame, B_j is obtained by assigning it the sum of BPA over all the compatible A_k.

Dempster's combination rule and the compatibility relationship are perhaps best explained by way of an example. Consider two dissimilar, but compatible, frames of discernment $\Theta_A = \{a_0, a_1, a_2\}$ and $\Theta_B = \{b_0, b_1, b_2\}$ with basic probability assignments:

$$m_1(A_1) = m_1(\{a_0, a_1, a_2\}) = 0.0 \qquad m_2(B_1) = m_2(\{b_0, b_1, b_2\}) = 0.2$$

$$m_1(A_2) = m_1(\{a_0, a_1\}) = 0.1 \qquad m_2(B_2) = m_2(\{b_0, b_1\}) = 0.4$$

$$m_1(A_3) = m_1(\{a_0, a_2\}) = 0.1 \qquad m_2(B_3) = m_2(\{b_0, b_2\}) = 0.3$$

$$m_1(A_4) = m_1(\{a_1, a_2\}) = 0.0 \qquad m_2(B_4) = m_2(\{b_1, b_2\}) = 0.0$$

$$m_1(A_5) = m_1(\{a_0\}) = 0.3 \qquad m_2(B_5) = m_2(\{b_0\}) = 0.1$$

$$m_1(A_6) = m_1(\{a_1\}) = 0.5 \qquad m_2(B_6) = m_2(\{b_1\}) = 0.0$$

$$m_1(A_7) = m_1(\{a_2\}) = 0.0 \qquad m_2(B_7) = m_2(\{b_2\}) = 0.0$$

Further, suppose that the compatibility relation of $A \rightarrow B$ is given by:
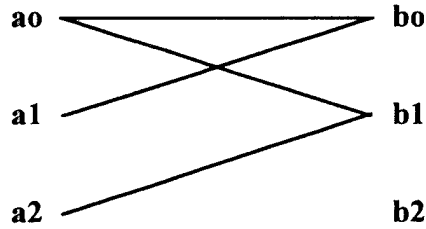
That is, the compatibility relation for the singletons is given by:

$$C_{A \rightarrow B}\left(A_7 = \{a_2\}\right) = \{b_1\}$$
$$C_{A \rightarrow B}\left(A_6 = \{a_1\}\right) = \{b_0\}$$
$$C_{A \rightarrow B}\left(A_5 = \{a_0\}\right) = \{b_0, b_1\}$$

Because the compatibility relation of a union of sets equals the union of the compatible subsets,

$$C_{A \rightarrow B}\left(A_i \cup A_j\right) = C_{A \rightarrow B}\left(A_i\right) \cup C_{A \rightarrow B}\left(A_j\right)$$

it follows that

$$C_{A \rightarrow B}\left(A_4 = \{a_1, a_2\}\right) = C_{A \rightarrow B}\left(A_7 \cup A_6\right) = \{b_0, b_1\},$$

$$C_{A \rightarrow B}\left(A_3 = \{a_0, a_2\}\right) = \{b_0, b_1\},$$

$$C_{A \rightarrow B}\left(A_2 = \{a_0, a_1\}\right) = \{b_0, b_1\}, \text{ and}$$

$$C_{A \rightarrow B}\left(A_1 = \{a_0, a_1, a_2\}\right) = \{b_0, b_1\}.$$

The compatibility relationship is illustrated in table 1 below.

Table 1. Compatibility Relationship

	$\{b_0\}$	$\{b_1\}$	$\{b_2\}$	$\{b_0,b_1\}$	$\{b_0,b_2\}$	$\{b_1,b_2\}$	$\{b_0,b_1,b_2\}$
$\{a_0\}$				x			
$\{a_1\}$	x						
$\{a_2\}$		x					
$\{a_0,a_1\}$				x			
$\{a_0,a_2\}$				x			
$\{a_1,a_2\}$				x			
$\{a_0,a_1,a_2\}$				x			

From the table, it is seen that the basic probability assignments in the frame θ_B from the first body of evidence are:

$$m_1\left(\{b_0,b_1,b_2\}\right) = 0.0,$$

$$m_1\left(\{b_0,b_1\}\right) = m_1(\{a_0\}) + m_1(\{a_0,a_1\}) + m_1(\{a_0,a_2\}) + m_1(\{a_1,a_2\}) + m_1(\{a_0,a_1,a_2\}) = 0.5,$$

$$m_1\left(\{b_0,b_2\}\right) = 0.0,$$

$$m_1\left(\{b_1,b_2\}\right) = 0.0,$$

$$m_1\left(\{b_0\}\right) = m_1\left(\{a_1\}\right) = 0.5,$$

$$m_1\left(\{b_1\}\right) = m_1\left(\{a_2\}\right) = 0.0,$$

$$m_1\left(\{b_2\}\right) = 0.0,$$

where the subscript m_i has been retained to differentiate the first body of evidence from the second. Now that the two bodies of evidence are represented in the same frame, Dempster's combination rule can be used to combine the two independent frames to a single *solution* frame. Dempster's combination rule is illustrated in table 2 below where the elements with zero belief are not shown.

Table 2. Dempster's Combination Rule

$m_1\backslash m_2$	$m_2(\theta B) = 0.2$	$m_2(\{b_0b_1\}) = 0.4$	$m_2(\{b_0b_2\}) = 0.3$	$m_2(\{b_0\}) = 0.1$
$m_1(\{b_0,b_1\}) = 0.5$	$b_0b_1\backslash 0.1$	$b_0b_1\backslash 0.2$	$b_0\backslash 0.15$	$b_0\backslash 0.05$
$m_1(\{b_0\}) = 0.5$	$b_0\backslash 0.1$	$b_0\backslash 0.2$	$b_0\backslash 0.15$	$b_0\backslash 0.05$

Here, the rows correspond to the elements of the power set of the first frame while the columns correspond to the second frame. Each cell in the table is associated with a member of the power

set in the combined frame given by the intersection of the sets associated with the corresponding row and column, and the belief is the product of the BPA of the individual sets. Only the non-zero entries are shown. The entries in each cell are the elements of the power set and the resulting portion of BPA to be assigned to that element. As there are no null intersections, $\alpha = 1$ by equation (6). Using equation (5), the belief values in table 2 are combined resulting in:

$$m_c(\{b_0,b_1\}) = 0.3,$$

$$m_c(\{b_0\}) = 0.7.$$

Using equations (3) and (4), the support for b_0 is 0.7, the plausibility is 1.0, and the uncertainty is 0.3, while the support for $\{b_0,b_1\}$ is equal to the plausibility for $\{b_0,b_1\}$, which is 1.0.

With a network of compatibility relations, different frames of reasoning can be linked together and combined. The collection of frames and compatibility relations and the place where they are combined is called a gallery. It is noted that differing galleries can be used for the same problem, and that in general each gallery can behave differently than another.

2.2 GEOMETRIC REPRESENTATION OF DS THEORY

Consider the generic frame of discernment comprising the set of N mutually exclusive and exhaustive hypothesis, θ_A, as given in equation (1) and repeated here for convenience as equation (9),

$$\Theta_A = \{a_1,...,a_N\}. \tag{9}$$

As previously noted, there are three useful measures of the elements of the power set A_i: the BPA, support, and plausibility. Although the interrelation of these three quantities is often complex for large frames, some insight may be obtained with a geometric representation of these measures.

Let each element of the power set be represented by a point in the three dimensional (space of real numbers R^3 with coordinates of plausibility, support, and belief, denoted by

$$f(A_i) = [p(A_i),\ s(A_i),\ m(A_i)]. \tag{10}$$

Because it is necessarily true that

$$p(A_i) \geq s(A_i) \geq m(A_i), \tag{11}$$

the space of allowable points that represents the elements of the power set is restricted to the interior and surface of the solid (right) triangle or wedge of unit height shown in figure 4.

Referring to figure 4, the following discussion illustrates the relationship between the distribution of BPA in the frame and the measures of support and plausibility. The terminology

introduced has been adopted to facilitate discussion and should not be considered as a standard. The point $(1,1,1)$ is the *certainty point*. That is, any element A_i that is represented by this point has all the belief assigned directly to it, and therefore, also has a support and plausibility of one. If BPA is redistributed from the element itself to its proper subsets, then its BPA is reduced while both plausibility and support remain at a value of one. Its representational point moves down the belief line $((1,1,0)-(1,1,1))$ until all the belief has been redistributed exclusively to its proper subsets. In that case

CERTAINTY POINT
(1,1,1)

BELIEF LINE

MAXIMUM
PLAUSIBILITY
PLANE

(1,1,0)
INCLUSION POINT

SUPPORT LINE   PLANE OF
NO BELIEF

(0,0,0)   PLAUSIBILITY LINE   (1,0,0)
EXCLUSION POINT   INDISMISSIBLE POINT
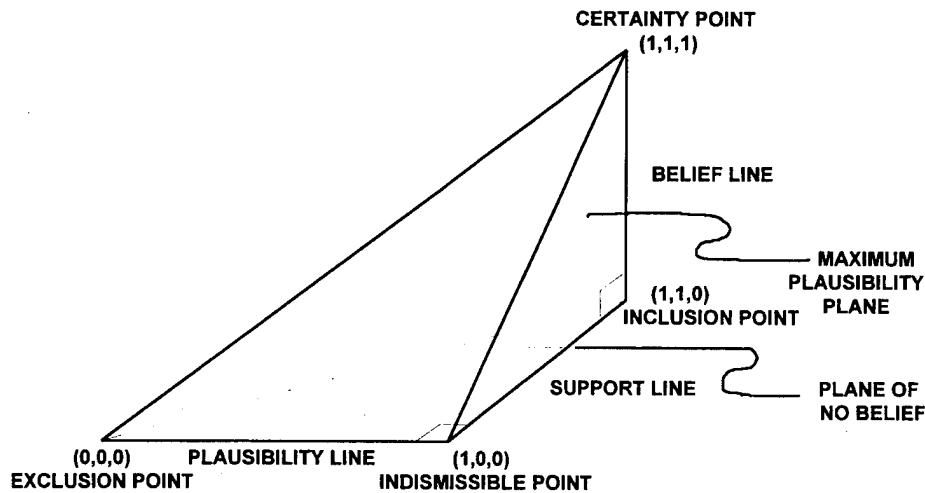
Figure 4. Geometric View of Dempster-Shafer Frame

there is no belief in the element itself and the coordinate location is the point $(1,1,0)$. This point has been labeled the *inclusion point* because although it has zero BPA, its subsets include all the available support and, therefore, must in theory include the correct hypothesis. If BPA is now redistributed from the proper subsets into sets that are not subsets but have non-empty intersections with A_i, then support for A_i is reduced while its plausibility remains one. The point moves along the support line $((1,0,0)-(1,1,0))$ until the point $(1,0,0)$ is reached. This point has been labeled the *indismissible point*. Although no evidence supports the hypotheses of this element, there is no evidence to the contrary. Thus, the plausibility is still high and there is no reason to dismiss these hypotheses. Further redistribution of BPA into the compliment of A_i and its subsets reduces plausibility. The resulting movement is along the plausibility line $((0,0,0)-(1,0,0))$ until the point $(0,0,0)$ is reached. At this point, all the belief is distributed among the subsets in the complement of A_i. This point has been labeled as the *exclusion point* because all available BPA is in the compliment, thus these hypotheses may be excluded. Other redistributions of belief to the subsets, supersets, intersecting sets, and compliment sets result in points located within or on the surface of the wedge. Note that more than one subset of the frame may be located at a given point in the solid triangle. However, there are some restrictions such as only $(n-1)$ elements of the power set may be located above the level of belief with value $(1/n)$.

2.3 HYPOTHESIS SELECTION CRITERION FOR DS THEORY

This section provides the details of the proposed hypothesis selection process for the DS frame of ER. Referring to figure 4, the proposed method projects all the points into the plane of maximum plausibility, or the support-belief plane. A belief cut is defined as a plane of constant belief, which becomes a line of constant belief when projected into the support-belief plane. Starting with the set(s) with the highest value of BPA, the union of all subsets represented by points on the belief cut is formed and the support of the resulting union computed. The process is repeated at the next highest level of belief, again taking the union of all the sets with BPA equal to or greater than the belief cut value and computing its support. The process is continued until some minimum threshold in support is met or exceeded. The final union contains all the hypotheses for further consideration. Although this approach may appear to be involved, this is not necessarily the case, since only those elements of the power set with non-zero BPA and their unions need to be enumerated. This can often yield a significant computational savings. Note that although the method results in an unambiguous selection, it is possible that other elements of the power set have the same support, however, the selected set will be a proper subset of these other elements.

The hypothesis selection method is equivalent to ranking the elements of the power set by BPA, then forming the union of subsets in order of decreasing BPA until a threshold on support of the union is met or exceeded. Specifically, let B be the ordered set

$$B = \{B_1,...,B_K\}, \tag{12}$$

such that

$$m(B_i) \geq m(B_j) \; ; \; j > i; \tag{13}$$

and where K is the number of elements in the power set as previously defined. Let m_d, $d = 1,2,...$ be the distinct values of belief where $m_d > m_{d-1}$, and n_d the number of sets with belief m_d.

Let the number of sets from each of the q belief cuts be L_q, where

$$L_q = \sum_{d=1}^{q} n_d , \quad q = 1,2,.... \tag{14}$$

Define the union of sets of B's as C_j

$$C_j = \bigcup_{i=1}^{j} B_i . \tag{15}$$

The selection criterion is then simply the set C_{Lq} with the smallest value of q, such that its support meets or exceeds the threshold λ,

$$\min_{L_q} s(C_{L_q}) \geq \lambda. \tag{16}$$

10

Since DS representation in computer software is most efficiently implemented by tracking only those subsets of the frame with non-zero belief, no additional elements need to be represented beyond those resulting strictly from application of the gallery, other than possibly a few that result from the union operation given in equation 15. Further, since the sets are combined in order of decreasing belief, support for the union of the sets grows in the most efficient manner.

To evaluate the selection criterion, a limited set of experimental results were obtained for a generic frame with four elements. The specific elements with non-zero BPA, as well as the number of elements with non-zero BPA, and their values, were selected from a uniform distribution. The power set is enumerated in table 3 with a selected set of BPA and the resulting measures of support and plausibility. The entries have been rounded for simplicity.

Table 3. Frame of Discernment

	BPA	Support	Plausibility
{a0,a1,a2,a3}	0	1	1
{a1,a2,a3}	0	0.80	1
{a0,a2,a3}	0	0.40	0.97
{a0,a1,a3}	0	0.88	0.88
{a0,a1,a2}	0	0.35	0.72
{a2,a3}	0	0.40	0.77
{a1,a3}	0.37	0.68	0.88
{a1,a2}	0	0.15	0.72
{a0,a3}	0	0.28	0.85
{a0,a2}	0	0.12	0.32
{a0,a1}	0.20	0.23	0.60
{a3}	0.28	0.28	0.65
{a2}	0.12	0.12	0.12
{a1}	0.03	0.03	0.60
{a0}	0	0	0.20

Arranging the subsets of the frame in order of decreasing (and non-zero) BPA produces the set ranking, {a1,a3}, {a3}, {a0,a1}, {a2}, and {a1} with BPA of 0.37, 0.28, 0.20, 0.12, and 0.03, respectively. Applying the selection algorithm, the union formed at the successive level cuts are {a1,a3} with support 0.68, {a0,a1,a3} with support 0.88, and {a0,a1,a2,a3} with support 1.0. If the support threshold is set above the value 0.68 but below 0.88, then the selected subset of hypotheses would be {a0,a1,a3}. For the example here only five sets have non-zero BPA and, in combination with their unions, require that only two out of a total of 15 support values must be computed for a the given support threshold.

Several other selection criteria for the subsets of the DS frame that involved distance measures were examined. One such measure is the distance of a subset from the certainty point (1,1,1). In this method subsets are collected (unioned) until the accumulated belief exceeds some

threshold. Because the frame Θ_A is located on the belief line $((1,1,0)-(1,1,1))$, it is always included once the radius has reached a value of one. Consequently, if the belief threshold has not been reached before the frame is encountered, a complex set of logic must be invoked. In addition to the logic problem, all the elements in the power set must be represented, which as noted previously, can be a large number for frames of modest size. Another method examined the distribution of points when projected down into the plane of no belief or the plausibility-support plane. Unfortunately, this technique is subject to the same limitations just mentioned.

3. CONTACT MANAGEMENT MODEL ASSESSMENT

As stated previously, the basic problem in contact management model assessment is to determine the set of models that are consistent with the observed data and to obtain the corresponding contact state estimates along with regions of uncertainty. The specific problem being considered here is the case of a change in the dynamic process from a known process. In this class of problems, a single correct model has been identified and used to obtain a good contact state estimate. At a subsequent point in time, the dynamic process or environment changes. The task is to determine which subset of models may be consistent with the new data without actually processing the data to obtain a state estimate. Primarily, this is accomplished by examining the predicted measurement residual sequences for deterministic features that are evidence of modeling errors. DS theory of evidential reasoning is used to determine which of the model hypotheses could possibly have produced the observed features. Section 3.1 describes the particular CMMA scenario under consideration. The application of DS evidential reasoning process to CMMA is given in section 3.2. The details of the functions used to generate and map evidence from the predicted residuals into belief in the frame of discernment are discussed in greater detail in sections 3.3 and 3.4.

3.1 CMMA APPLICATION

The CMMA application considered here is the tracking of a single dynamic contact by an observer with data from two spatially diverse acoustic sensors in an ocean environment. One sensor provides measurements of azimuthal bearing (β_A), depression angle (DE), and frequency. The azimuthal bearing is the angle of signal arrival in the horizontal plane and the DE is the angle of signal arrival in the vertical plane. The other sensor is a line array that provides measurements of conical angle (β_c) and frequency. The conical angle is the angle of signal arrival relative to the array axis and is a function of both the azimuthal bearing and the depression angle.

In one approach to contact management tracking, the data are segmented by a statistical hypothesis test and compressed to an approximate sufficient statistic. The compressed parameters are then subsequently processed to obtain a contact-state estimate.[16,17] Usually the data are segmented according to own ship legs. However, sometimes the cause for a segmentation in the data is a change in the kinematic, sensor, or environmental structure of the process. Thus, there may be a need to test the data for consistency with the model, prior to processing. For the class of problems considered here, the data prior to the time of a change in the structure of the problem are processed using the correct models to obtain a maximum likelihood estimate of the contact's state. A subsequent segment of data, possibly produced by a different process and, thus, constituting a change in the model, is assessed for a subset of consistent models to be selected for further processing. A total of three model changes are used to define this CMMA application. The models are a change in propagation path, a change in base frequency, and a change in the contact velocity vector (maneuver). The three model changes, along with the original model (null hypothesis) and a category for an unclassified model anomaly, yield a total of five model hypotheses from which a selection must be made.

Each of the predicted residual sequences of the five measurements that are available from the two sensors are examined for anomalous features such as a jump in the data sequence, a drift, or both. The presence of these features in the predicted residual sequences are evidence that there has been a change in the model. For example, a change in the propagation path from direct path to bottom bounce results in a jump in the depression and conical angles, while having no effect on the azimuthal bearing. The features are subsequently evaluated by the evidential reasoning system to determine which model(s) could (or could not) have produced the observed effects. The CMMA gallery that is representative of this reasoning process is discussed in greater detail in the following section.

3.2 CMMA GALLERY DESCRIPTION

An overview of the CMMA evidential reasoning system, or gallery, is shown in figure 5. In the first stage of the CMMA system, evidence is extracted from the available measurements. In this case, an existing state estimate and the underlying process model are used to obtain predicted residuals. The predicted residuals are examined for the deterministic features of jump and drift which are used to establish the basic probability assignments, $m(A_i)$, in the feature frame of discernment. One feature frame of discernment is established for each measurement type. A compatibility relation is used to map the BPA in the features for each measurement type into BPA in the elements of the model hypotheses frame of discernment. Again, there is one compatibility relation for each measurement type. Dempster's combination rule is used to combine the model hypotheses frames for all measurement types into a single solution frame of discernment. The selection criteria can then be applied to the solution frame to determine which modeling hypotheses are likely to be consistent with the evidence presented in the predicted residuals and, therefore, should be considered for further processing.
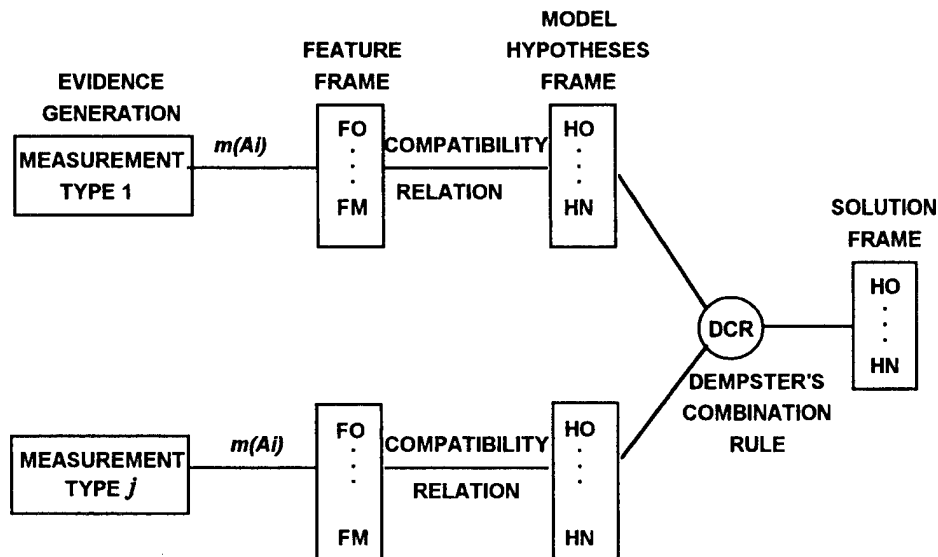

Figure 5. Block Diagram of CMMA System

14

A more detailed representation of the first stage in the CMMA gallery is given in figure 6. Here, the establishment of the feature frame has been broken down into its constituent parts to facilitate discussion.
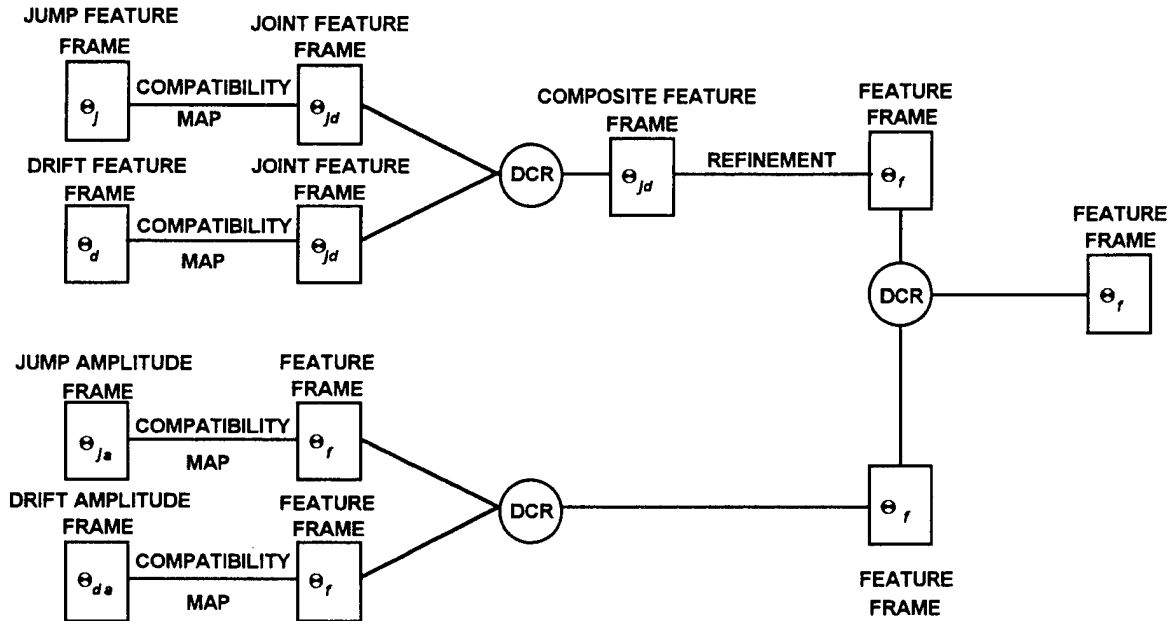
Figure 6. Complete Feature Frame Block Diagram

The gallery used to establish the feature frame of discernment is divided into two lines or paths of reasoning. The first line of reasoning involves evidence pertaining to the existence of the features and is depicted in the upper half of figure 6. The second line of reasoning involves evidence of the features' amplitudes and is shown in the lower half of figure 6. Since the initial frames of discernment established for each type of evidence are different but related, the intervening stages of the gallery map belief from the initial frames into compatible frames of discernment, which are then combined to produce the final or complete frame of discernment. This final feature frame of discernment is shown as the right most block in figure 6 and is the feature frame used in figure 5.

The first stage of the gallery comprises two feature existence frames and two feature amplitudes frames of discernment. The individual frames of discernment for the existence of a feature are shown for the features of jump, $\Theta_j = \{$jump (j), no-jump $(\bar{j})\}$, and drift, $\Theta_d = \{$drift (d), no-drift $(\bar{d})\}$, in the upper portion figure 6. For each measurement type, a traditional likelihood ratio test is used to generate evidence pertaining to the existence of the features of a step (jump) and/or a ramp (drift). The functions that map the evidence of a feature into BPA in the feature's existence in the frame of discernment are discussed in section 3.3.

15

The second stage in establishing the feature existence are the joint feature frames of discernment q_{jd} = {no features ($\bar{j}\ \bar{d}$), jump-only ($j\bar{d}$), drift-only ($\bar{j}d$), jump and drift (jd)}, again one for each measurement type. BPA in the joint feature frame is obtained from BPA in frames q_j and q_d through the compatibility relationship defined in table 4.

Table 4. Single Feature Frame To Joint Feature Frame
Compatibility Relationship

Single	Feature Frame	Joint Feature Frame
Θ_j	$\{\bar{j}\ \}$	$\{\bar{j}\ \bar{d}, \bar{j}\,d\}$
	$\{j\}$	$\{j\bar{d}, jd\}$
	$\{\bar{j}\,j\}$	$\{\bar{j}\ \bar{d}, \bar{j}\,d, j\bar{d}, jd\}$
Θ_d	$\{\bar{d}\ \}$	$\{\bar{j}\ \bar{d}, j\bar{d}\ \}$
	$\{d\}$	$\{\bar{j}\,d, jd\}$
	$\{\bar{d}, d\}$	$\{\bar{j}\ \bar{d}, \bar{j}\,d, j\bar{d}, jd\}$

Using the compatibility map of table 4, two joint feature frames are generated, one for the jump feature and one for the drift feature. From this table, it is seen that the existence of a jump feature in θ_j is compatible with the hypotheses of jump-only or jump and drift ($j\bar{d}$, jd) in power set of θ_{jd}. Consequently, the BPA for jump from θ_j is mapped into BPA for the element ($j\bar{d}$, jd) of θ_{jd}. Similar reasoning applies to the remaining terms in the compatibility relationship. Note that BPA for the elements of the individual feature frames are mapped to BPA in subsets of the joint feature frame, so BPA can be assigned directly into this frame. Although the system BPA's were implemented directly in the joint feature frame, the above description is useful in developing the evidence representations.

BPAs in the two joint feature frames are merged using Dempster's combination rule to form a single composite feature frame. Note that although the constituent joint feature frames have no BPA in the singleton hypotheses, e.g., $\{\bar{j}\ \}$ from θ_j and $\{d\}$ from θ_d map and combine to $\{\bar{j}\,d\}$ in θ_{jd}. The combined frame may have non-zero BPA for the singleton hypotheses. The composite feature frame of discernment is then refined to reflect amplitudes in the feature frame of discernment of weak, moderate, and strong. The feature frame, Θ_f is:

$$\Theta_f = \{\bar{j}\bar{d}, \bar{j}dw, \bar{j}dm, \bar{j}ds, jwdw, \cdots, jsdm, jsds\},$$

where $jwdw$ is weak jump and weak drift, etc. The refinement is implemented with the compatibility map defined in table 5. Here, for example, the BPA in a jump-only is compatible with the set of weak-jump and no drift, moderate-jump and no drift or strong-jump and no drift. The other relationships are similarly defined in table 5.

16

Table 5. Composite Feature Frame to Feature Frame Compatibility Relationship

Θ_{jd}	Θ_f
{ $\bar{jd}$ }	{ $\bar{j}\ \bar{d}$ }
{ $j\bar{d}$ }	{ $jw\bar{d}, jm\bar{d}, js\bar{d}$ }
{ $\bar{j}d$ }	{ $\bar{j}\,dw, \bar{j}\,dm, \bar{j}\,ds$ }
{ jd }	{ $jwdw, jmdw, jsdw, jwdm, \ldots, jsds$ }

The second line of reasoning in the gallery for establishing the feature frame of discernment is the feature amplitude frame. Here, the estimates of the jump or drift amplitude and their standard deviation, obtained from equation (22) in section 3.3, are used to establish BPA in the amplitude of the feature in the feature frame. The functions used to map the evidence into BPA are discussed in section 3.4. The frame of discernment is {weak feature, moderate feature, and strong feature}. There are separate frames for both the jump, Θ_{ja}, and drift, Θ_{da}, features. A compatibility map relates the BPA in feature amplitude to BPA in the feature frame and is given in table 6 for the jump feature. A similar compatibility relationship is used for the drift feature and the two frames are combined using Dempster's combination rule. Finally, the last step is combining the belief in the feature frame established by the evidence for feature existence with that from feature amplitude to produce the final feature frame shown as the right most block of figure 6 and the second block of figure 5.

Table 6. Jump Amplitude to Feature Frame Compatibility Relationship

Jump Amplitude Frame	Feature Frame of Discernment
$\{js\}$	$\left\{ js\bar{d}, jsdw, jsdm, jsds \right\} = A_1$
$\{jm\}$	$\left\{ jm\bar{d}, jmdw, jmdm, jmds \right\} = A_2$
$\{jw\}$	$\left\{ jw\bar{d}, jwdw, jwdm, jwds \right\} = A_3$
$\{jm,js\}$	$A_2 \cup A_1$
$\{jw,js\}$	$A_3 \cup A_1$
$\{jw,jm\}$	$A_3 \cup A_2$
$\{jw,jm,js\} = \Theta_{ja}$	$A_3 \cup A_2 \cup A_1$

Referring to the overall CMMA block diagram of figure 5 again, the feature frames for each of the residual types are mapped into a model hypothesis frame using a compatibility map. The model hypotheses frame of discernment comprises the model events: no change (Ho), change in base frequency (BF), change in propagation path (PP), contact maneuver (CM), and unknown anomaly (NL). There is one compatibility map and resultant model hypothesis frame for each

17

residual type. The compatibility map represents the logical relationship between the occurrence of a feature set in a residual sequence and the corresponding model hypotheses that could be the cause. For example, a change in propagation path results in a weak, moderate, or strong jump in the depression angle with a respective drift amplitude of no drift, weak, or moderate. Thus, the events $\{jw\overline{d}\}, \{jmdw\}, \{jsdm\}$ in the depression angle are all compatible with a change in the propagation path. Consequently, the belief in all the subsets of $\{jw\overline{d}, jmdw, jsdm\}$ in DE are compatible with the model hypothesis of change in propagation path. Thus, their BPAs are assigned to the model hypothesis of change in propagation path. If a subset of the feature frame is compatible with more than one model hypothesis, its belief is associated with the union of the model hypotheses. The compatibility relationships were determined using a perturbation analysis, and were verified experimentally via scatter diagrams. There are over 90 relationships in the compatibility map. The compatibility relationships are detailed in appendix A and the perturbation analysis is detailed in appendix E.

The final solution is obtained by combining all the beliefs in all the model hypothesis frames using Dempster's combination rule. At this point, the selection criterion can be applied to select modeling hypotheses for further processing of the data as indicated in figure 2. Although this CMMA application considered limited amounts of evidence and model hypotheses, the framework of this gallery is readily extendible to include additional evidence from features such as curvature in the predicted residual sequence, additional evidence obtained directly from the measurements or other sources, as well as additional model hypotheses.

3.3 FEATURE EXISTENCE EVIDENCE REPRESENTATIONS

The evidence representations map the evidence pertaining to the existence of features in the predicted residuals into belief in the feature frame of discernment. Although these are not evidence representations as defined by Shafer,[12] the usage of that term has pervaded many of the references and is, therefore, retained here. Note, however, that the correct term is "a function that assigns basic probability."

As mentioned in the preceding section, the belief in the feature frame is established in two stages. In the first stage, belief in the existence of the feature is established using a test statistic from a multiple hypothesis test. The test statistic is an estimate of the signal to noise ratio of the various features. It is computed as the squared difference between two regression fits to the residuals, one with all the coefficients (features), the other with some coefficients set to zero.[13] The belief in a feature is based on the difference between the probability of detection (PD) and probability of false alarm (PF), with the threshold set at the point of equal likelihood of the two densities associated with PD and PF. The belief in no feature is established as the joint probability of two (assumed) independent events: the probability of false alarm, and the belief in the ability to detect a reference level feature in the given noise level. The approach of the first stage is summarized in this section and the details are given in appendix C and by Ferkinhoff, et al.[13] The second stage is a refinement of the belief in a feature into belief in a particular amplitude for the feature, that is, weak, moderate, or strong jump and/or drift. These beliefs are established by computing the expected value of the membership functions for each of the amplitudes: weak, moderate, or strong. The second stage is detailed in Appendix D and by Ferkinhoff.[14]

3.3.1 EVIDENCE EXTRACTION

Evidence of features in the predicted residual sequence is obtained in the form of the estimated coefficients of a polynomial regression fit. Only the pertinent results are presented in this section. Additional detail may be found in appendix B. Consider the following discrete time dynamic state model that generates measurements, $z(k)$:

$$x(k + 1) = \Phi x(k) + Bu(k) + \Gamma w(k), \tag{17a}$$

$$z(k) = H[x(k)] + n(k). \tag{17b}$$

Here, $x(k)$ is the $nx1$-dimensional state vector, u is the input vector, w is zero-mean, white, Gaussian process noise, z is the measurement, and n is zero-mean, white, Gaussian measurement noise, which is independent of the process noise. The matrices Φ, B, Γ are, respectively, the state transition matrix, input matrix, and process noise input matrix. The function $H[*]$ is the nonlinear function relating the state to the measurement and k is the time index. The state $x(k)$ is available only as an estimate, $\hat{x}(k|k)$, along with the estimated error covariance matrix, $P(k|k)$. The estimate $\hat{x}(k|k)$ of the state at time k is based on all the past data up to and including time k. The estimate is available from a suitable estimator such as an extended Kalman filter or maximum likelihood estimator. The predicted residual $\tilde{z}(k|k-1)$ is the difference between the measurement $z(k)$ and the estimate of the measurement, $\hat{z}(k|k-1)$, that is,

$$\tilde{z}(k|k-1) = z(k) - \hat{z}(k|k-1), \tag{18a}$$

where

$$\hat{z}(k|k-1) = H[\hat{x}(k|k-1)], \tag{18b}$$

$$\hat{x}(k|k-1) = F\hat{x}(k-1|k-1) + Bu(k-1), \tag{18c}$$

and where $\hat{x}(k|k-1)$ is the state estimate at time k based on the measurements up to and including time $k-1$ but not time k. The process to be investigated here is the predicted residual sequence,

$$\tilde{Z} = [\tilde{z}_1, \tilde{z}_2, \ldots \tilde{z}_N]^T, \tag{19a}$$

where

$$\tilde{z}_i = \tilde{z}(k + i|k), \tag{19b}$$

$$\widetilde{z}(k + i|k) = z(k + i) - H\,[\hat{x}(k + i|k)],\tag{19c}$$

$$\hat{x}(k + i|k)] = F^i\,\hat{x}(k|k) + \sum_{j=1}^{i} F^{(j-1)}\,Bu(k + j - 1).\tag{19d}$$

Here, as well as in appendix B, only a single measurement type is considered at a time. The predicted residual sequence can be approximated as a jointly Gaussian random vector. However, the approximation only remains valid whenever the second-order and higher terms of the Taylor series expansion of the measurement nonlinearity can be ignored. Thus, two assumptions on the predicted residual sequence are invoked to assure the reasonableness of the Taylor series approximation. First, the dynamic model of equation (17), which is used to obtain the estimate, must be a reasonably accurate description of the real process. Second, the state error, and its effect over the time interval of interest, must remain sufficiently small to allow the validity of a first-order Taylor series approximation of the predicted residual sequence. These two assumptions are the basis of the null hypothesis (H_0: no modeling error and a good state estimate).

Under the H_0 hypothesis, the predicted residual process is a zero-mean jointly Gaussian random vector; that is, the probability density function is

$$P(\widetilde{Z}|H0) = \left[\sqrt{2p}^{\,N} det\,|V\,|^{1/2}\right]^{-1} exp\left[-\frac{1}{2}\;\widetilde{Z}^T V^{-1}\widetilde{Z}\right]\tag{20a}$$

where V is the covariance matrix given by,

$$V = E\left[\widetilde{Z}\widetilde{Z}^T\right] - E\left[\widetilde{Z}\right]E\left[\widetilde{Z}^T\right].$$

Note that although elements of the predicted residual vector are correlated, there always exists a set of basis vectors in which $\widetilde{Z}$ can be represented by uncorrelated (and hence independent) components. For the applications considered here, both process noise and state estimation error covariances are small compared with the measurement noise covariance. For these cases the elements of $\widetilde{Z}$ are also independent

Under the alternate hypotheses, $H_j, j \neq 0$, there is a modeling error. The original state and measurement models of equations (17) do not adequately describe the system dynamics. The modeling errors may arise from unknown system inputs, a change in the measurement model, or an unobserved change in the measurement noise statistics. Examples of such modeling anomalies are, respectively, a target maneuver, change in propagation path, and measurement bias. Regardless of the cause of the modeling error, its observable effect is often a feature in the predicted residual sequence.

Under the mismodeling hypotheses $H_j, j \neq 0$, the predicted residuals exhibit features that are distinguishable from those of the hypothesis H_0. For the hypothesis $H_j, j \neq 0$, an (j- 1)-the order

20

polynomial is used to model the anomalous features of the predicted residuals. The probability density function for the jointly Gaussian random vector under H_j, $j \neq 0$, is

$$P(\tilde{Z}|H_j) = \left[\sqrt{2p}^N \ det \ |V|^{1/2} \right]^{-1} \exp \left[-\frac{1}{2} \ (\tilde{Z} - M_1)^T V^{-1} (\tilde{Z} - M_1) \right], j \neq 0 \tag{21a}$$

where,

$$M_1 = B\overline{m} , \tag{21b}$$

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & (t_1 - t_0) & (t_1 - t_0)^2 & \cdots \\ & \cdot & \\ & \cdot & \\ & \cdot & \end{bmatrix}, \tag{21c}$$

and

$$\overline{m} = \begin{bmatrix} m_0 \\ m_1 \\ \cdot \\ \cdot \\ \cdot \\ m_{l-1} \end{bmatrix}. \tag{21d}$$

The features of interest are extracted from a regression fit of a second-order polynomial to the residuals referenced to time t_0:

$$\overline{m} = \left[B^T B \right]^{-1} B^T \left[Z - Z_0(\hat{x}) \right], \tag{22}$$

where $Z_0(\hat{x})$ are the predicted measurements associated with hypothesis H_0, the B matrix is given in equation (21c), and $\overline{m} = \left[\overline{m}_0, \overline{m}_1, \overline{m}_2 \right]^T$ is the vector of estimated regression coefficients identified with the specific features of jump (m_0), drift (m_1), and curvature (m_2). Although curvature and higher order terms may be present in the residual sequence, they are not used presently as evidence, however, their presence may be necessary in the regression fit to obtain unbiased estimates of the jump and drift coefficients. Here, the significance of the curvature coefficient is rejected if its test statistic (discussed below) falls below a threshold set to achieve a 10-percent false alarm rate. Other techniques are also available to determine the appropriate order of the regression fit.[18]

An example of a noise-free residual plot is shown in figure 7. Here, β_A β_C have been displayed. The figure displays the difference between the measurement and its estimate on the horizontal axis with time progressing downward on the vertical axis. For the purposes of illustration, there is essentially no noise on the measurements. Data up to and including the time line marker labeled t_r have been used to obtain the state estimate. Data subsequent to the time line are the predicted residuals. Here, the process changed from direct path propagation to bottom bounce. The predicted residual plot for conical angle exhibits a jump and a slight drift; the azimuthal bearing exhibits no features. The following sections explain how this evidence may be used to limit the class of models to those that can produce these effects.

CONICAL ANGLE   **AZMUTHAL BEARING**

TIME LINE t_r
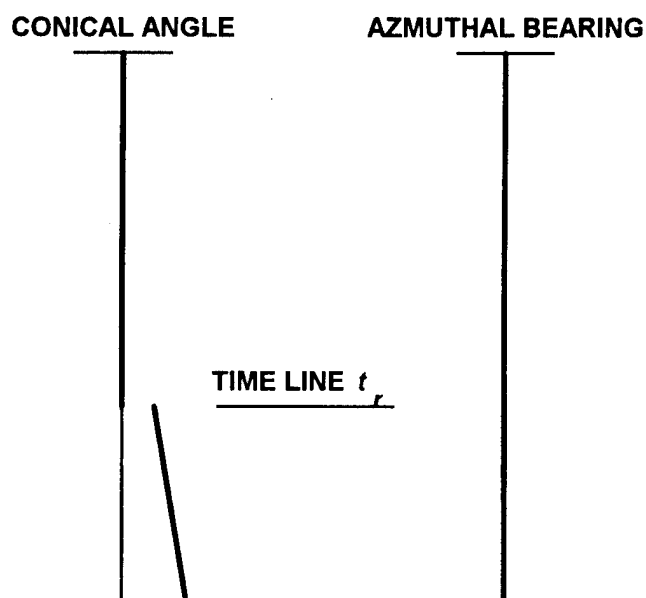
Figure 7. Noise-Free Residuals Plot

3.3.2 EVIDENCE REPRESENTATION

The basis for distributing belief are the feature detection and false alarm probabilities, P_D and P_F obtained from a multiple hypothesis likelihood ratio test. The test statistic l is compared with a threshold ε for the multiple hypothesis test as

$$l \underset{\bar{H}_j}{\overset{\bar{H}_i}{\gtrless}} \varepsilon , \qquad (23)$$

where H_j is the hypothesis that all features exist, H_i is the hypothesis that the ith feature is not significant, and

$$l = \frac{\left\| \hat{M}_i - \hat{M}_j \right\|^2}{\sigma^2} . \qquad (24)$$

22

Here, $\hat{M}_k = B\hat{\bar{m}}_k$, where the $\hat{\bar{m}}_k$ are separate estimates with the appropriate coefficients set to zero, and the test statistic l is an estimate of the signal(feature)-to-noise ratio (SNR). Traditionally, the threshold is set to satisfy a specified probability of false alarm criterion.

The test statistic is the squared difference between the two regression fits, $\hat{M}_j$ with all the coefficients and $\hat{M}_i$ with some coefficients set to zero. With Gaussian distributed errors and known variance the test statistic is a chi-squared random variable with the number of degrees of freedom equal to the number of coefficients being tested, one in this case. Since the test statistic is chi-squared with one degree of freedom, it may be alternatively viewed as the square of a univariant Gaussian random variable.[19] Under the hypothesis that a feature is not present, the corresponding Gaussian distribution is zero mean. Under the hypothesis that a feature is present, the Gaussian distribution has a non-zero mean equal to the square root of the test statistic. In the absence of any prior knowledge about the feature, the threshold is taken at the point of equal likelihood of the two densities, which is at $\varepsilon = \sqrt{l} / 2$. Note that the threshold varies with the estimated SNR and, therefore, responds dynamically to changing signal or noise levels. While the assumption of known variance is used here, unknown variance can also be handled by substituting an F-test for the chi-squared test.

The P_F is associated with the belief that the estimated SNR value belongs to the "no feature" distribution (true SNR = 0), while P_D is associated with the belief that the estimated SNR value belongs to the "feature present" distribution (true SNR $\neq$ 0). The belief that a feature is present is directly proportional to the estimated SNR, which is analogous to a signal excess (excess beyond SNR = 0). Consequently, the belief that a feature, i.e. jump, is present is defined as the difference between P_D and P_F:

$$m(j) = P_D - P_F. \tag{25}$$

The belief in no feature (no-jump) $m(\bar{j})$ is more complicated. Two separate conditions lead to zero SNR. In one case, noise may approach infinity; in the other, the signal level may approach zero. In the large noise case, although SNR is approaching zero, the placing of all belief into the no feature hypothesis results in low plausibility of a feature being present. This is clearly undesirable, because although high-noise inhibits a feature detection, one could be present; thus the plausibility of the existence of a feature should be high. For the simple structure of the individual feature frames, a high plausibility for the feature implies a low belief in no feature. Thus, $m(\bar{j})$ should be a function of both the estimated SNR as well as a measure of the significance of the noise level. Following this line of reasoning, $m(\bar{j})$ is established as the joint probability of two (assumed) independent events: the probability of false alarm P_F given the estimated SNR, and the belief in the ability to detect a reference level feature in the given noise level. Consequently

$$m(\bar{j}) = (P_{DR} - P_{FR}) P_F, \tag{26}$$

where P_{DR} and P_{FR} are, respectively, the detection and false alarm probabilities of a reference signal in the given noise level. The reference signals can be set at the minimum level of bias that

can be reasonably tolerated. The importance of the reference signal level is discussed in the experimental results section. Because $m(\bar{j}) \leq P_F$, it follows that

$$m(\bar{j}) + m(j) \leq P_D \leq 1, \tag{27}$$

and, since the BPAs must sum to unity, the remaining belief is assigned to the frame and represents ignorance as

$$m(\Theta_j) = 1 - P_D + [1 - (P_{DR} - P_{FR})]P_F . \tag{28}$$

A more detailed discussion of the functions that map the evidence of features in the predicted residuals into the basic probability assignments of the feature existence frames is given by Ferkinhoff, et al.[10,14]

3.4 FEATURE AMPLITUDE EVIDENCE REPRESENTATIONS

The second line of reasoning involved evidence of the features' amplitude, that is, weak, moderate, or strong jump or drift. Evidence of feature amplitude is embedded in the probability density function of the estimated features, or regression coefficients of equation 22. The feature density function is approximated with a Gaussian density function where the estimated features are used as the mean values and the covariance is obtained from equation 22. The functions that map evidence of weak, moderate, and strong features into the basic probability assignments in the feature amplitude frame of discernment are illustrated graphically below in figures 8a through 8c and are discussed in greater detail in appendix D and by Ferkinhoff, et al.[14]

As an example, the probability density function for a feature, jump, is shown in figure 8a. Figure 8b illustrates the regions defined as weak, moderate, and strong jumps with polarity discrimination (±). Note that in this construction, the boundaries of the membership functions can overlap, thus allowing for a fuzzy set interpretation for feature amplitudes. The expected values, for each jump amplitude; strong negative (-js), moderate negative (-jm), weak negative or positive (jw), moderate positive (+jm), and strong positive (+js), are obtained as the integration of the product of the feature's probability density and membership functions, and are illustrated in the histogram of figure 8c. The BPA is assigned from this histogram. Level cuts are made at each value as shown in the figure with dashed lines. These level cuts are projected to the scale on the far right. The BPA is the difference in these values times the number of sets intersected in the histogram. The resulting BPA value is assigned to the union of the sets that have been intersected by the area between dashed lines in the plot. However, because the amplitude frame does not currently distinguish between positive and negative features, the final BPA assignments are without regard to polarity. The compatibility relation between frames with and without polarity are shown in table 7.

p (JUMP)
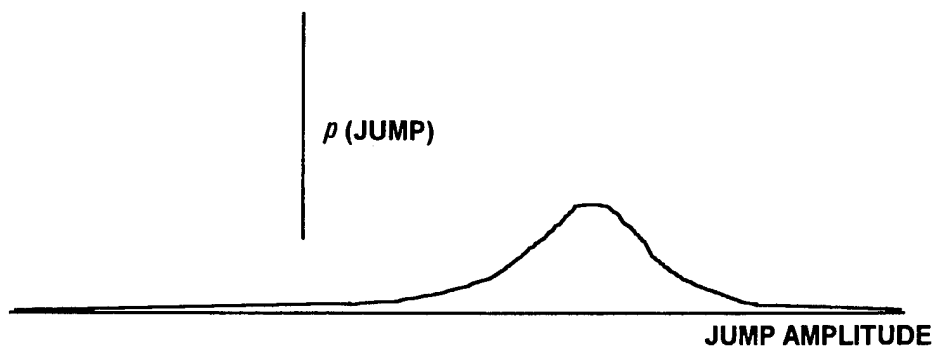
JUMP AMPLITUDE

Figure 8a.   Feature Amplitude Probability Density Function

STRONG-   MODERATE-   **1.0**   MODERATE+   STRONG+

WEAK

a   b   c   d   e   f   g   h   JUMP AMPLITUDE
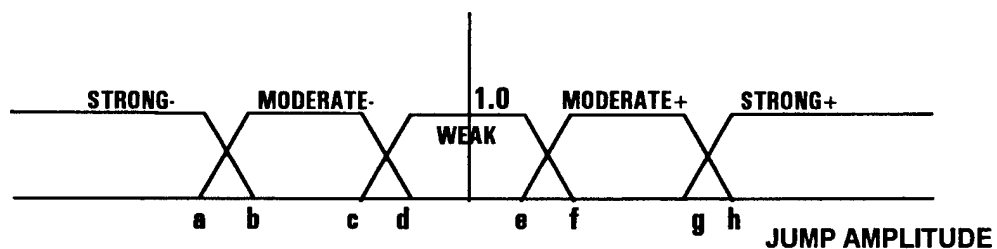
Figure 8b.   Feature Amplitude Membership Functions

e5 *m(+jm)=e5-e4*

e4 *m(jw,+jm) = 2(e4-e3)*
e3 *m(jw,+jm,+js) = 3(e3-e2)*

e2 *m(jw,jm,+js) = 3(e2-e1)*
e1 *m(jw,jm,js) = 5(e1)*

E{js}   E{-jm}   E{jw}   E{+jm}   E{+js}

Figure 8c.   Feature Amplitude Expected Value and BPA

25

Table 7. Compatibility Relation Between Feature Amplitudes With and Without Polarity

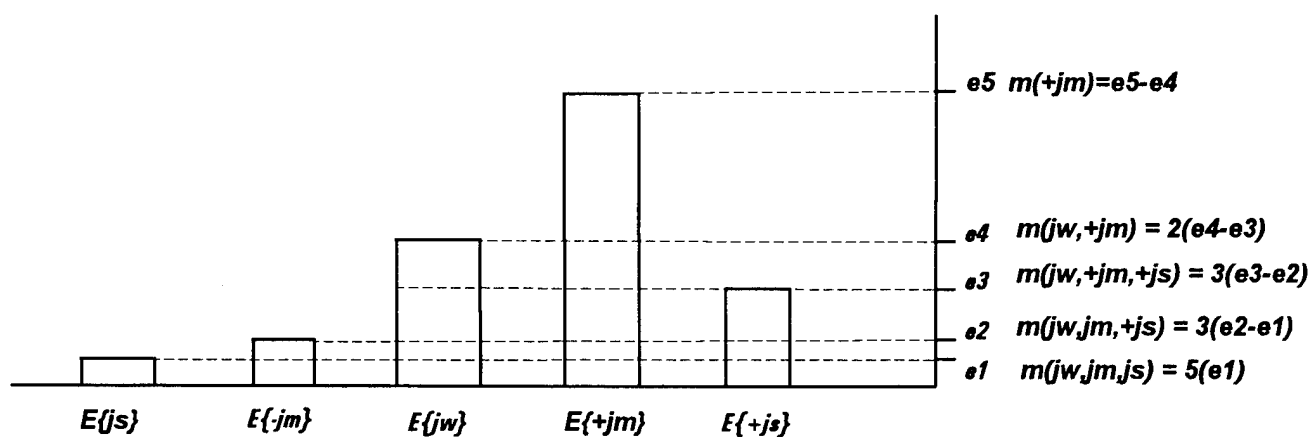| Feature Amplitude Frame With Polarity	Feature Amplitude Frame Without Polarity
$\{jw,-jm,+jm,-js,+js\}$	$\{jw,jm,js\}$
$\{jw,-jm,+jm,-js\}$	$\{jw,jm,js\}$
$\{jw,-jm,+jm,,+js\}$	$\{jw,jm,js\}$
$\{jw,-jm,-js\}$	$\{jw,jm,js\}$
$\{jw,+jm,+js\}$	$\{jw,jm,js\}$
$\{jw,-jm,+jm,\}$	$\{jw,jm\}$
$\{jw,-jm\}$	$\{jw,jm\}$
$\{jw,+jm\}$	$\{jw,jm\}$
$\{-jm,-js\}$	$\{jm,js\}$
$\{+jm,+js\}$	$\{jm,js\}$
$\{+js\}$	$\{js\}$
$\{-js\}$	$\{js\}$
$\{-jm\}$	$\{jm\}$
$\{+jm\}$	$\{jm\}$
$\{jw\}$	$\{jw\}$

Note that for single mode probability density functions, like the Gaussian, logically inconsistent assignments of BPA to sets like $\{-js,+js\}$ or $\{jw,-js\}$ cannot occur with this technique. Consequently, non-zero BPA for both sets $\{-js\}$ and $\{+js\}$ cannot occur simultaneously, however, a non-zero BPA assignment for the sets $\{-js\}$ and $\{-js,-jm,jw,+jm,+js\}$ is possible and does not create a logically inconsistent result. Also, because the membership functions that define the regions of feature amplitudes sum to unity for all values of jump amplitude in figure 8b, a total BPA of unity is obtained in the frame of discernment.

The boundaries for the feature amplitude regions were obtained from a computer simulated perturbation analysis. The perturbation analysis was performed numerically with a monte carlo simulation and the results presented in the form of scatter plots of event occurrence in a jump-drift feature plane. The scatter plots are discussed in appendix E. The boundaries were established by visual inspection; no attempt was made to optimize the boundary locations.

4. EXPERIMENTAL RESULTS

Experimental results were obtained using synthetic data generated by a computer simulation. As described in the previous section, the CMMA simulation involved a single contact and a single observation platform with two spatially diverse sensors. The available measurements are: azimuthal bearing, depression angle, and frequency from the first sensor, and conical angle and frequency from the second sensor. The measurement noise standard deviation were arbitrarily selected as 0.3°, 0.5°, 0.03, 0.7°, and 0.04 Hz, respectively. The data rate for all measurements are 20 seconds.

Initially, the contact is located 5000 m from the observer at a bearing of 10° and moves on a course of 330° with a speed of 12 m/s. The base frequency is a nominal 800 Hz. The observer executes a three-leg geometry, where each leg is a segment of constant velocity. For these experiments, the time on each leg is 900 s and the course and speed for the legs are 300° and 14 m/s, 60° and 14 m/s, and 340° and 12 m/s, respectively.

The data from the first two ownship legs are processed using the correct model hypotheses. The segment of data from the third leg is examined for model consistency and the evidential reasoning process is used to select a subset of model hypotheses for further processing. For the purposes of these experiments, there is only a single change in the model which is known to occur in the middle of the third leg. The set of model changes is a change in propagation path from direct path to bottom/surface reflection (PP), a contact maneuver (CM), a change in the base frequency (BF), no change in the model or the null hypothesis (H0), and an unknown model change (NL).

The experimental results are contained in table 8. Each experiment is represented by the entries in a column. The rows of the table are the set of hypotheses selected by the CMMA system over all the experiments. The entries in the table are the percent of time that the corresponding hypotheses set for the given experiment were selected. The selection threshold was set at the support level of 0.9. For each experiment the results were obtained from 100 monte carlo trials. Six different experiments were conducted, involving five different modeling anomalies and one that had no model change. The first experiment, shown in column one, was a small contact maneuver (SMCM) involving a course change of 5° to a new heading of 335°. The second experiment was a larger contact maneuver (LGCM) of 15° to a new course of 345°. The third experiment was a small change in the propagation path (SMPP) from direct path to bottom bounce with a shallow ocean bottom depth of 500 m. The fourth experiment was a large change in propagation path from direct path to bottom bounce with a bottom depth of 3000 m. The fifth experiment was a change in the base frequency (BF) of 15 Hz. The last experiment had no model error and is the null hypothesis (H0).

As shown by the results in table 8 for the SMCM experiment, the hypothesis set of change in base frequency, propagation path, or contact maneuver was selected most often (52 percent) while the sets base frequency or contact maneuver, and propagation path or contact maneuver were selected 25 percent and 22 percent, of the time respectively. The incorrect hypothesis base frequency was selected 1 percent of the time while contact maneuver alone was never selected.

Table 8. Experimental Results

	SMCM	LGCM	SMPP	LGPP	BF	H0
{H0,BF,PP,CM}	0	0	1	0	0	22
{BF,PP,CM}	52	0	1	0	0	1
{H0,BF,PP}	0	0	0	0	0	14
{H0,PP,CM}	0	0	27	0	0	14
{BF,CM}	25	0	0	0	0	0
{PP,CM}	22	0	70	0	0	0
{H0,PP}	0	0	0	0	0	1
{H0,BF}	0	0	0	0	0	33
{CM}	0	100	1	0	0	0
{PP}	0	0	0	100	0	0
{BF}	1	0	0	0	100	2
{H0}	0	0	0	0	0	13

Since this contact maneuver is small and, therefore, of limited observable consequence, it is not surprising that the correct modeling hypothesis, contact maneuver, is never uniquely identified. More importantly, however, is that the correct hypothesis is included 99 percent of the time. Thus, for this example, the evidential reasoning is not very decisive but is inclusive. Compare the results of the small contact maneuver in column one with the results for the larger contact maneuver shown in column two; here, the correct hypothesis is always selected. Clearly the larger contact maneuver, which has a more observable effect on the residuals, leads to more decisive reasoning.

Similar results are also observed for the small and large changes in propagation path given in columns three and four. Note however, that for small changes in the propagation path, although the correct hypothesis is included 99 percent of the time, it is associated with the null hypothesis 28 percent of the time . That is, 27 percent of the time the evidential reasoning process concludes that either a change in propagation path, contact maneuver, or nothing happened and 1 percent of the time concludes {H0,BF,PP,CM}. When changes in the model are associated with H0, a conclusion that may be drawn is that whatever may have occurred is of such a small effect as to be confused with H0. Therefore, one approach is to ignore the possible model changes and select H0 until further evidence warrants a different conclusion. This confusion with H0 is apparent in the null hypothesis experimental results given in the last column. Here the correct hypothesis, H0 is included 97 percent of the time while the false detection rate is 3 percent. Note that H0 alone is selected only 13 percent of the time. Currently, it is believed that this behavior is attributable to the manner in which the evidence is generated and represented for CMMA and is being investigated further.

The results presented are typical of those obtained from many simulations and have been selected as representative. However, extensive analysis under conditions of poor observability with limited data were not attempted. Thus, although current results indicate that performance degrades gracefully with deteriorating conditions, a complete analysis of that subject has not been conducted.

5. CONCLUSION

The application of DS theory of evidential reasoning to the contact management model assessment problem was examined. This report is a collection of results that have been reported separately elsewhere but are presented together here for the first time. Several results were obtained. Evidence representations, that map evidence from statistical hypothesis test to the belief in the existence of features is developed

A method for hypothesis selection in the DS framework of evidential reasoning was developed. The method involves arranging the subsets of the frame in order of decreasing BPA values. For each distinct value of BPA all the subsets with BPA equal to or exceeding that value are unioned, and the support for the unioned set determined. The process is continued with decreasing BPA values until a threshold on the support is met or exceeded. The resulting set is the set of selected hypotheses. The selection method was applied to the evidential reasoning processes of the contact management model assessment problem and experimental results were obtained with synthetic data. Although limited results were obtained, the selection criterion achieved acceptable performance in that the selected set included the correct model hypothesis. The selection method is efficient in that, at most, it only requires the computations of support for the elements of the power set with non-zero BPA and their unions. A geometric representation for the frame based on BPA values and the measures of support and plausibility was also presented and has provided insight into the evidential reasoning process.

Evidence representations were developed that map the evidence obtained from a statistical hypothesis test to belief in the features of jump and drift. The belief is established in two stages. In the first stage, belief in the existence of the feature is established using a test statistic from a multiple hypothesis test. The test statistic is an estimate of the signal to noise ratio of the various features. It is computed as the squared difference between two regression fits to the residuals, one with all the coefficients (features), the other with some coefficients set to zero. The belief in a feature is based on the difference between the probability of detection (PD) and probability of false alarm (PF), with the threshold set at the point of equal likelihood of the two densities associated with PD and PF. The belief in no feature is established as the joint probability of two (assumed) independent events: the probability of false alarm, and the belief in the ability to detect a reference level feature in the given noise level. The second stage is a refinement of the belief in a feature into belief in a particular amplitude for the feature, that is , weak, moderate, or strong jump or drift (or both). These beliefs are established by computing the expected value of the membership functions for each of the amplitudes. Beliefs from both stages are then combined using Dempster's combination rule.

Experimental results, using simulated data, were presented that demonstrated the viability of the DS theory of evidential reasoning to the contact management model assessment problem. In almost all cases (99 percent) the correct model hypothesis was included in the selected hypothesis set. For scenarios with strongly observable features the correct hypothesis was almost always the single hypothesis selected from the universal set. The case for no model change presented some confused results in that it was seldom the single selection, however it almost always included in the selected set. The cause for this confusion is believed to be the result of

not having a membership set, in the feature amplitude frame of discernment, to indicate that the detected feature was a very small feature of no consequence.

Suggestions for future work are to expand the sources of evidence to include the processed residuals under the selected models, evidence in the measurements, and solution quality regions for the multiple hypothesis estimates. Further investigation into the use of fuzzy sets, fuzzy rule based reasoning,[20] and the use of possibility theory are recommended. Additionally, Bayesian inference nets may also be investigated.

30

APPENDIX A
COMPATIBILITY RELATIONS

The CMMA compatibility relationships for the five measurements of sphere bearing, sphere depression angle, sphere and towed array frequency, and conical angle are given in tables A-1 through A-4. Due to the large frame size only, the compatibility relations between the singletons of both frames are given. Since the compatibility of a union equals the union of the compatible subsets, all remaining relationships between the supersets of the singletons may be readily derived. These relationships were obtained using both an analytical and numerical perturbation analysis. The results of the numerical analysis are contained in appendix E.

Referring to table A-1, the feature frame has 16 elements,

$$\theta_f = \{\, \bar{j}\ \bar{d}\ , jw\bar{d}\ , jm\bar{d}\ , js\bar{d}\ , \bar{j}\ dw, \bar{j}\ dm, \bar{j}\ ds, jwdw, jwdm, jwds, jmdw, jmdm, jmds, jsdw, jsdm, jsds\}$$

which are weak, moderate, and strong jump with no drift ($jw\bar{d}\ , jm\bar{d}\ , js\bar{d}$), respectively, weak, moderate, strong drift with no jump ($\bar{j}\ dw, \bar{j}\ dm, \bar{j}\ ds$), respectively; and all their pairwise combinations. The elements of the model frame

$$\theta_M = \{H0, BF, PP, CM, NL\}$$

are: no model change ($H0$), change in base frequency (BF), change in propagation path (PP), contact maneuver (CM), and an unidentifiable model change (NL). The x's entered in the table signify compatible events. For the sphere bearing compatibility relations given in table A-1, no feature ($\bar{j}\ \bar{d}$) is compatible with no model change, a change in base frequency, or a change in propagation path, thus the mapping is

$$\{\bar{j}\ \bar{d}\ \} \rightarrow \{H0, BF, PP\}$$

A similar interpretation can be given to the remaining entries in all the tables.

Table A-1. Compatibility Relations for Sphere Bearing

	Ho	BF	PP	CM	NL
$\bar{j}\,\bar{d}$	x	x	x		
$jw\bar{d}$				x	
$jm\bar{d}$				x	
$js\bar{d}$					x
$\bar{j}\,dw$				x	
$\bar{j}\,dm$				x	
$\bar{j}\,ds$					x
$jwdw$				x	
$jwdm$				x	
$jwds$					x
$jmdw$				x	
$jmdm$				x	
$jmds$					x
$jsdw$					x
$jsdm$					x
$jsds$					x

Table A-2. Compatibility Relations for Sphere Depression Angle

	Ho	BF	PP	CM	NL
$\bar{j}\,\bar{d}$	x	x		x	
$jw\,\bar{d}$			x	x	
$jm\,\bar{d}$			x		
$js\,\bar{d}$					x
$\bar{j}\,dw$				x	
$\bar{j}\,dm$				x	
$\bar{j}\,ds$					x
$jwdw$			x	x	
$jwdm$				x	
$jwds$					x
$jmdw$			x		
$jmdm$			x		
$jmds$					x
$jsdw$					x
$jsdm$					x
$jsds$					x

Table A-3. Compatibility Relations for Sphere and Towed Array Frequency

	Ho	BF	PP	CM	NL
$\bar{j}\,\bar{d}$	x				
$jw\,\bar{d}$		x	x	x	
$jm\,\bar{d}$		x		x	
$js\,\bar{d}$		x			
$\bar{j}\,dw$			x	x	
$\bar{j}\,dm$				x	
$\bar{j}\,ds$					x
$jwdw$		x	x	x	
$jwdm$				x	
$jwds$					x
$jmdw$		x	x		
$jmdm$			x		
$jmds$					x
$jsdw$		x			
$jsdm$					x
$jsds$					x

Table A-4. Compatibility Relations for Towed Array Conical Angle

	Ho	BF	PP	CM	NL
$\bar{j}\,\bar{d}$	x	x			
$jw\,\bar{d}$			x	x	
$jm\,\bar{d}$			x		
$js\,\bar{d}$					x
$\bar{j}\,dw$			x	x	
$\bar{j}\,dm$				x	
$\bar{j}\,ds$					x
$jwdw$			x	x	
$jwdm$				x	
$jwds$					x
$jmdw$			x		
$jmdm$			x		
$jmds$					x
$jsdw$					x
$jsdm$					x
$jsds$					x

APPENDIX B
PROCESS MODEL

Consider the following discrete, time-dynamic state model that generates measurements, $z(k)$:

$$x(k + 1) = Fx(k) + Bu(k) + Gw(k), \tag{B-1a}$$

$$z(k) = H[x(k)] + n(k). \tag{B-1b}$$

Here, $x(k)$ is the $nx1$ dimensional state vector, u is the input vector, w is zero-mean, white, Gaussian measurement noise, z is the measurement, and n is zero-mean, white, Gaussian measurement noise, which is independent of the process noise. The matrices F, B, G are, respectively, the state transition matrix, input matrix, and process noise input matrix. The function H is the nonlinear function relating the state to the measurement and k is the time index.

The state $x(k)$ is available only as an estimate, $\hat{x}(k|k)$, along with the estimated error covariance matrix, $P(k|k)$. The estimate $\hat{x}(k|k)$ of the state at time k is based on all the past data up to and including time k. The estimate is available from a suitable estimator like an extended Kalman filter or maximum likelihood estimator. The residual $\tilde{z}(k|k-1)$ is the difference between the measurement $z(k)$ and the estimate of the measurement, $\hat{z}(k|k-1)$, that is,

$$\tilde{z}(k|k-1) = z(k) - \hat{z}(k|k-1), \tag{B-2a}$$

where

$$\hat{z}(k|k-1) = H[\hat{x}(k|k-1)], \tag{B-2b}$$

$$\hat{x}(k|k-1) = \Phi\hat{x}(k-1|k-1) + Bu(k-1), \tag{B-2c}$$

and where $\hat{x}(k|k-1)$ is the state estimate at time k based on the measurements up to and including time $k-1$ but not time k. The process to be investigated here is the predicted residual sequence,

$$\tilde{Z} = [\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_N]^T, \tag{B-3a}$$

where

$$\tilde{z}_i = \tilde{z}(k+i|k), \tag{B-3b}$$

$$\tilde{z}(k+i|k) = z(k+1) - H[\hat{x}(k+i|k)], \tag{B-3c}$$

$$\hat{x}(k+i|j) = \Phi^i \hat{x}(k|k) + \sum_{j+1}^{i} \Phi^{(j-1)} Bu(k+j-1). \tag{B-3d}$$

Examining the ith element of the predicted residual sequence

$$\tilde{z}(k + i|k) = H[x(k + i)] - H[\hat{x}(k + i|k)] + n(k + i), \tag{B-4}$$

and expanding the measurement function in a Taylor series about the estimate, with

$$\delta x(k + i) = \hat{x}(k + i|k) - x(k + i), \tag{B-5}$$

yields the approximate expression

$$\tilde{z}(k + i|k) = -a^{\mathrm{T}}[\hat{x}(k + i|k)]\delta x(k + i) + n(k + i), \tag{B-6a}$$

where

$$a[\hat{x}(k + i|k)] = \frac{\partial Hx(k+i)}{\partial x(k+i)} \Big|_{x(K+i) = \hat{x}(k+i)} . \tag{B-6b}$$

Noting that the dynamic equation for the propagation of the state estimation error in terms of the fixed error at time k is

$$\delta x(k + i) = \Phi^i \delta x(k) - g^T(k + i)W, \tag{B-7a}$$

where

$$g^T(k + i) = [\Phi^{i-1}\Gamma, \Phi^{i-2}\Gamma, ..., \Phi\Gamma, \Gamma, 0, ..., 0], \tag{B-7b}$$

and

$$W^T = [w(0), w(1), ..., w(i-2), w(i - 1), ..., w(N-1)]. \tag{B-7c}$$

The entire predicted residual sequence can be approximated as

$$\tilde{Z} = -A\big[\hat{x}(k|k)\big]\delta x(k) - GW + N_n, , \tag{B-8a}$$

where

B-2

$$A [\hat{x}(k|k)] = \begin{bmatrix} \vdots \\ \left(\dfrac{\partial H[x(k+i)]}{\partial x(k)}\right)^{t} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ a^{T}[\hat{x}(k+i|k)]F^{i} \\ \vdots \end{bmatrix}_{x(k)=\hat{x}(k|k)} \tag{B-8b}$$

and

$$N_n = \begin{bmatrix} \vdots \\ n(k+i) \\ \vdots \end{bmatrix}, G = \begin{bmatrix} \vdots \\ g^{T}(k+1) \\ \vdots \end{bmatrix} \tag{B-8c}$$

The predicted residual sequence can be approximated as a jointly Gaussian random vector. For unbiased state estimates, $\widetilde{Z}$ is zero mean and its probability density function is

$$P(\widetilde{Z}) = \left[\sqrt{2\pi}^{\,n} \, det \; |V|^{1/2} \right]^{-1} exp \left[-\frac{1}{2} \widetilde{Z}^{T} V^{-1} \widetilde{Z} \right]. \tag{B-9a}$$

Here, V is the covariance matrix of the predicted residual vector defined as

$$cov(\widetilde{Z}) = E\left[\widetilde{Z}\widetilde{Z}^{T}\right] - E\left[\widetilde{Z}\right]E\left[\widetilde{Z}^{T}\right], \tag{B-9b}$$

$$cov(\widetilde{Z}) = E[(N_n - A\delta x)(N_n - A\delta x)^{T}]. \tag{B-9c}$$

Thus,

$$V = AP(k|k)A^{T} + S^{2} + GQG^{T} \tag{B-9d}$$

where

$$S^2 = E[N_n N_n^T],$$

(B-9e)

$$P(k|k) = E[dx(k)dx^T(k)],$$

(B-9f)

$$Q = E[WW^T].$$

(B-9g)

Note that the presence of process noise in the predicated residual sequence results in an effective noise,

$$N_e = N_n - GW$$

(B-10a)

with an increased covariance,

$$S_e = S^2 + GQG^T.$$

(B-10b)

Further, it is noted that the probability density function for the predicted residuals is only approximately Gaussian. However, the approximation remains valid whenever the second-order and higher terms of the Taylor series expansion of the measurement nonlinearity can be ignored. Thus, two assumptions on the predicted residual sequence are invoked to assure the reasonableness of the Taylor series approximation. First, the dynamic model of equation (1), which is used to obtain the estimate, must be a reasonably accurate description of the real process. Second, the state error, and its effect over the time interval of interest, must remain sufficiently small to allow the validity of a first-order Taylor series approximation of the predicted residual sequence. The two assumptions stated above are the basis of the null hypothesis, H_0: No modeling error and a good state estimate.

Under the H_0 hypothesis, the predicted residual process is a zero-mean jointly Gaussian random vector; that is, the probability density function is

$$P\left(\tilde{Z}|H_0\right) = \left[\sqrt{2\pi}^N \; det\left|V\right|^{1/2}\right]^{-1} exp\left[-\tfrac{1}{2}\tilde{Z}^T V^{-1}\tilde{Z}\right],$$

(B-11)

where V is the covariance matrix given by equation (B-9d). Note that although elements of the predicted residual vector are correlated, there always exists a set of basis vectors in which $\tilde{Z}$ can be represented by uncorrelated (and, therefore, independent) components. For many applications considered here, both process noise and state estimation error covariances are small compared with the measurement noise covariance. For these cases, the elements of $\tilde{Z}$ are approximately independent

Under the alternate, or other hypothesis, H_1, there is a modeling error. The original state and measurement models of equations (B-1) do not adequately describe the system dynamics. As mentioned in the introduction, modeling errors may arise from unknown system inputs, a change in the measurement model, or an unobserved change in the measurement noise. Examples of such

B-4

modeling anomalies are, respectively, a target maneuver, change in propagation path, and measurement bias. Regardless of the cause of the modeling error, its observable effect is often a feature in the predicted residual sequence. Under hypothesis H_0 the predicted residual process is approximately zero-mean jointly Gaussian, and for small state error with no process noise, is a white process.

Under the mismodeling hypothesis H_1, the predicted residuals exhibit features that are distinguishable from those of the hypothesis H_0. For the hypothesis H_1 an (l-1)-th order polynomial is used to model the anomalous features of the predicted residuals, that is

$$H_1: \tilde{Z}(k+i|k) = -a^T\left[\hat{x}(k+i|k)\right]\delta\hat{x}(k) + m(k+i) + n(k+i) - g^T(k+i)W, \quad \text{(B-12a)}$$

where

$$m(k+i) = [a_0 + a_1(t_{k+i} - t_0) + a_2(t_{k+i} - t_0)^2 + ,..., + a_{l-1}(t_{k+i} - t_0)^{l-1}]u(t_{k+i} - t_0), \quad \text{(B-12b)}$$

$$u[t] = \begin{cases} 0 & t < 1 \\ 1 & t > 1 \end{cases}, \text{ and } t_0 \text{ is the time of the modeling anomaly.}$$

The predicted residual sequence under H_1, is

$$H1: \tilde{Z} = -A[\hat{x}(k|k)]\delta\hat{x}(k) - GW + M_1 + N_n \quad \text{(B-13a)}$$

where

$$M_1 = B\bar{m} \quad \text{(B-13b)}$$

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & (t_1 - t_0) & (t_1 - t_0)^2 & \cdots \\ & \cdot & \\ & \cdot & \\ & \cdot & \end{bmatrix}, \quad \text{(B-13d)}$$

and

$$\bar{m} = \begin{bmatrix} a_o \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_{l-1} \end{bmatrix} \qquad\qquad\qquad (\text{B-13c}).$$

The additive term, due to mismodeling, M_1, contributes to the mean value of $\tilde{Z}$, hence

$$E[\tilde{Z}] = M_1 \qquad\qquad\qquad (\text{B-14a})$$

and

$$cov\, E[\tilde{Z}] = E[(\tilde{Z} - M_1)(\tilde{Z} - M_1)^T] = V. \qquad\qquad (\text{B-14b})$$

Thus, the covariance under both assumptions is the same and only the means differ. The probability density function for the jointly Gaussian random vector under H_1 is

$$P(\tilde{Z}|H_1) = \left[\sqrt{2\pi}^N \, det|V|^{1/2} \right]^{-1} exp\left[-\tfrac{1}{2}(\tilde{Z} - M_1)^T V^{-1}(\tilde{Z} - M_1) \right] \qquad (\text{B-15})$$

Under hypothesis H_1 (H_1: A modeling anomaly exists) the probability density function of the predicted residuals is (possibly) non-zero mean jointly Gaussian,

$$p(\tilde{Z}|H_1) \, Þ \, N(M_1, V) \qquad\qquad\qquad (\text{B-16})$$

where M_1 is the mean vector defined in equation (B-13) and V is the covariance matrix given by equation (B-9d), and $N(M_1, V)$ is the Gaussian probability density function with Mean M_1, and covariance V.

Similar results can be developed for both the smoothed residuals and the filtered innovations. For the smoothed residual process the state estimation error is correlated to the past measurement noise. Noting that the conditional probability density function or likelihood function for the observations (no process noise)

$$Z = [z_1,...,z_n]^T, \qquad\qquad\qquad (\text{B-17})$$

given the state x is,

$$P(Z|x) \Rightarrow N(Z(x), S^2), \qquad\qquad\qquad (\text{B-18a})$$

B-6

or

$$P(Z|x) = \left[\sqrt{2p}^N det|S^2|^{1/2}\right]^{-1} exp\left[-\frac{1}{2}(Z - Z(x))^T S^{-2}(Z - Z(x))\right],$$ (B-18b)

where

$$Z(x) = [H[x(1)],...,H[x(N)]].$$ (B-18c)

The maximum likelihood estimate is the value of the state, x, that maximizes the likelihood function of equation (B-18). Taking the logarithm of the likelihood function yields the log-likelihood equation.

$$ln[p(Z|x)] = [constant] + \frac{1}{2}(Z - Z(x))^T S^{-2}(Z - Z(x)).$$ (B-19)

Differentiating the log-likelihood equation with respect to the state at time k and setting the result to zero, the maximum likelihood estimate can be obtained by solving the equation

$$A^T[x(k)]S^{-2}[Z - Z(x(k))]\Big|_{x(k)=\hat{x}(k)} = 0 \quad .$$ (B-20)

For the linearized smoothed residuals

$$\tilde{Z}_s = -A[\hat{x}(k|k)]dx(k) + N_n$$ (B-21)

where $A[x(k|k)]$ is defined in equation (B-8). The error in the estimate, $dx(k)$ is approximately

$$dx(k) \gg [A^T S^{-2} A]^{-1} A^T S^{-2} N_{n,,}$$ (B-22a)

and the state error covariance is

$$P(k|k) = E\{(dx(k)dx^T(k))\},$$

(B-22b)

which can be approximated as

$$P(k|k) \gg [A^T S^{-2} A]^{-1}.$$ (B-22c)

Consequently, from equation (B-9), the probability density function for the smoothed residuals under H_0 is

$$P(\tilde{Z}|H_0) \triangleright N(0 V_s),$$ (B-23a)

where

$$V_s = cov(\widetilde{Z})$$

or

$$V_s = E[(N_n - Adx)(N_n - Adx)^T];$$

and, substituting from equation (B-22) for $dx(k)$ yields

$$V_s = s^2 + AP(k|k)A^T. \tag{B-23b}$$

Similarly, under the hypothesis H_1

$$P(\widetilde{Z}_s|H_1) \triangleright N(M_1, V_S). \tag{B-24}$$

When the filtered residual sequence is obtained from the extended Kalman filter, the estimate $x(k/k)$ is the conditional mean, and the filtered residuals or innovations process is white.[21] For this process, the probability density function is

$$H_0: P(\widetilde{Z}_f|H_0) \triangleright N(0, V_f) \tag{B-25a}$$

$$H_1: P(\widetilde{Z}_f|H_1) \Rightarrow N(M_1, V_f) \tag{B-25b}$$

where

$$V_f = diag[a^T P(i|i)a + s_n^2]. \tag{B-25c}$$

The above development characterizes three processes. They are, predicted residuals, smoothed residuals, and filtered (conditional mean) residuals. The predicted residuals are obtained from future measurements and the state prediction, while the smoothed residuals are obtained by fitting the past data with the best state estimate. The filtered residuals, usually obtained from a Kalman filter, involve the current state estimate and the next measurement. While the filtered innovation or residuals obtained from a Kalman filter are a white process, both the smoothed and predicted residuals are correlated to some degree; however, for small state estimation error they can often be approximated as a white process. In this investigation, attention is focused on the predicted residuals, and for some results is limited to the case of negligible state estimation error.

B-8

APPENDIX C
MULTIPLE HYPOTHESIS TEST

In the previous section, the problem of detecting a signal with known structure in the presence of noise was considered. Since that test considers one of two possible outcomes, either the signal is present or it is not, it is often called a binary hypothesis test. This section discusses the multiple hypothesis tests.

The anomalous feature model was defined by equation (B-12b) as an lth order polynomial. In the binary hypothesis test of the previous section the signal detection problem was considered. In that case the signal structure was known, that is, polynomial order was known as well as which coefficients are non-zero. Here, only the maximum model order is known and a test is devised to determine which polynomial coefficients are non-zero. This is a standard problem in regression analysis and the development presented here is adapted from Stuart et al. and Hoel et al.[22,23]

The case considered here is a maximum polynomial order of one for the anomalous feature; that is

$$\bar{m} = [a_0, a_1]^T. \tag{C-1}$$

The hypotheses are:

H_0: $[a_0 = 0, a_1 = 0]$ no anomaly present - noise only,
H_1: $[a_0 \neq 0, a_1 = 0]$ offset or jump anomaly only,
H_2: $[a_0 = 0, a_1 \neq 0]$ drift - only anomaly, and
H_3: $[a_0 \neq 0, a_1 \neq 0]$ jump and drift anomaly

The development is easily extended to arbitrary model order by noting that the set of ℓ polynomial coefficients results in 2^{ℓ} hypotheses.

Consider the special case of uncorrelated error in the predicted residual sequence. This corresponds to the case of small or no state estimation error. In this case the model for the predicted residual of equation (18-a) of the text is

$$\tilde{Z} = N_n + M. \tag{C-2}$$

Since the measurement noise vector N_n is independent, zero mean, Gaussian random vector, it follows that

$$P(\tilde{Z}) \Rightarrow N[M, s_n^2 I].$$

The joint density function of the predicted residuals, conditioned on the hypothesis H_i that the

*i*th anomalous feature is present in the sequence (i.e., jump-only, drift-only, jump-drift, etc.) is

$$P(\widetilde{Z}|H_i) \triangleright N[M_i, s_n^2 \, I \,,$$

where M_i is the corresponding mean vector induced by the anomaly of the *i*th hypothesis. The likelihood ratio test between any two hypothesis, H_i and H_j, is

$$LRT(\widetilde{Z}) = \frac{P(\widetilde{Z}|H_i)}{P(\widetilde{Z}|H_j)} \overset{\overline{H}j}{\underset{\overline{H}i}{\underset{<}{>}}} 1. \qquad \text{(C-3)}$$

Here, if the *LRT* exceeds the threshold *l*, then the hypothesis, H_j, corresponding to the presence of the *j*th feature in $\widetilde{Z}$, is rejected. Likewise, if the *LRT* is below threshold, then hypothesis H_i is rejected.

The case where the measurement noise standard deviation, s_n, is known, and the anomaly, $M = [m_1,...,m_N]^T$, is unknown is considered first. Substituting for $p(\widetilde{Z}|H)$ for the general case results in

$$LRT(\widetilde{Z}) = \frac{\left[(2\pi)^{N/2} det \left|V\right|^{1/2}\right]^{-1} exp\left[-\frac{1}{2}\left(\widetilde{Z}-M_i\right)^T V^{-1}\left(\widetilde{Z}-M_i\right)\right]}{\left[(2\pi)^{N/2} det \left|V\right|^{1/2}\right]^{-1} exp\left[-\frac{1}{2}\left(\widetilde{Z}-M_j\right)^T V^{-1}\left(\widetilde{Z}-M_j\right)\right]} \overset{-H_j}{\underset{-H_i}{\underset{<}{>}}} 1. \qquad \text{(C-4)}$$

Taking the natural logarithm yields

$$1N\left[LRT(\widetilde{Z})\right] = \frac{1}{2}\left\|\widetilde{Z}-M_i\right\|_{V^{-1}}^2 + \frac{1}{2}\left\|\widetilde{Z}-M_j\right\|_{V^{-1}}^2 \overset{\overline{H}j}{\underset{\overline{H}i}{\underset{<}{>}}} ln\lambda. \qquad \text{(C-5)}$$

Since *M* is unknown its maximum likelihood estimate is used to obtain the generalized likelihood ratio test (*GLRT*). The maximum likelihood estimate of $\overline{m}$ is

$$\hat{\overline{m}} = \left[B^T V^{-1} B\right]^{-1} B^T V^{-1} \widetilde{Z} \,, \qquad \text{(C-6a)}$$

and the estimate of predicted residual mean *M*, is

$$\hat{M} = B\hat{\overline{m}} \,. \qquad \text{(C-6b)}$$

C-2

For the case of interest, $V = s_n^2 I$ and since M is unknown, the estimate $\hat{M}$ of equation (C-6b) and (C-6a) are used in (C-5) yielding the generalized likelihood ratio test (*GLRT*)

$$\ln[GLRT(\tilde{Z})] = \frac{\left\|\tilde{Z} - \hat{M}_j\right\|^2}{\sigma_n^2} - \frac{\left\|\tilde{Z} - \hat{M}_i\right\|^2}{\sigma_n^2} \underset{\overline{Hi}}{\overset{\overline{Hj}}{\gtrless}} \ln\lambda^2 \qquad \text{(C-7a)}$$

or

$$\sum_{k=1}^{N} \frac{\left(\tilde{z}_k - \hat{m}_{jk}\right)^2}{\sigma_n^2} - \sum_{k=1}^{N} \frac{\left(\tilde{z}_i - \hat{m}_{ik}\right)^2}{\sigma_n^2} \underset{\overline{Hi}}{\overset{\overline{Hj}}{\gtrless}} \ln\lambda^2, \qquad \text{(C-7b)}$$

where $\hat{m}_{jk}$ is the kth component of the vector $\hat{M}_j$. Define the test statistic for the multiple hypothesis GLRT as

$$'_m(\tilde{Z}) = \ln\left[GLRT(\tilde{Z})\right] \qquad \text{(C-7c)}$$

To obtain quantifiable results for the GLRT, it is necessary to know the probability density function of the test statistic l_m. The remainder of this section is devoted to the development of the test statistic's probability density function.

From equation (C-2), the mean vector M can be decomposed into a linear combination of columns of the matrix B, that is

$$M = a_0, B_1 + a_1 B_2 + ,..., + a_{\ell-1} B_\ell \qquad \text{(C-8a)}$$

where B_i are the columns of B

$$B = [B_1| ... | B_l] \qquad \text{(C-8b)}$$

and

$$\tilde{m} = [a_0,...,a_{\ell-1}]^T. \qquad \text{(C-8c)}$$

The set of vectors $\{B_i\}_{i=1}^{\ell}$ are independent and form a basis of an ℓ-dimensional subspace of the N-dimensional space of the predicted residual vector. Let the jth hypothesis be the anomaly modeled by the ℓ-1 order polynomial with all coefficients present (H_3 for the case of interest). The other hypotheses are developed as constraints on the coefficients, that is

$$C_m^- = 0 \tag{C-9}$$

For the case of interest here, the hypothesis H_i result in the following constraint matrices, C,

$$H_0: C_0 = I_{2x2}; \qquad H_1: C_1 = [0,1]; \qquad H_2: C_2 = [1,0]$$

where I2x2 is the 2x2 identity matrix.

When hypothesis $H_i(i{\neq}j)$ is true, p-constraints are imposed on the coefficients, a_i, that can be used to reduce the number of independent terms in the decomposition of M in equation (C-9). Consequently, if $H_i = H_2$ then $a_1 = 0$,

$$M_1 = a_0 B_1, \tag{C-10}$$

where the rank of C equals p, and the space spanned by the remaining columns of B has been reduced by one. In general, if p-constraints are imposed by the ith hypothesis, then

$$M_i = a_0' n_i +,\dots,+ a_{\ell-p-1}' n_{\ell-p} \,, \tag{C-11}$$

where a_i' are the surviving coefficients and n_i are the basis vectors of (ℓ-p)-dimensional subspace spanned by M under H_i.

Let the set of nx1 vectors $\{r_i\}_{i=1}^{N}$ be an orthonormal basis for the M-dimensional space of the predicted residuals. Specifically, let the first (ℓ-p) vectors, $\{r_i\}_{i=1}^{\ell-p}$ be the basis for the space spanned by the constrained columns of B, and further, let the set of vectors $\{r_i\}_{i+1}^{\ell}$ span the column space of B, then

$$\widetilde{Z} = R\underline{a}, \tag{C-12a}$$

and
$$M = R\underline{g}, \tag{C-12b}$$

where,

$$R = [r_1| \dots |r_N] \tag{C-12c}$$

and $\underline{a}$ and $\underline{g}$ are (Nx1)-vectors of the coefficients of $\widetilde{Z}$ and M, respectively, with respect to the orthonormal basis vectors $\{r_i\}_{i=1}^{N}$ and

$$\underline{a} = [a_1,\dots,a_N]^T \tag{C-12d}$$

and

C-4

$$\gamma = [\gamma_1,\ldots,\gamma_{\ell\text{-}p}, 0, \ldots 0]^T. \tag{C-12e}$$

Substituting equation (C-12) into the two forms of equation (C-5), under the condition that the test hypothesis H_i is true, results in

$$\|\tilde{Z} - M_j\|^2 = \sum_{k=1}^{\ell}(a_k - g_k)^2 + \sum_{k=\ell+1}^{N}a_{k^2}, \tag{C-13a}$$

and

$$\|\tilde{Z} - M_i\|^2 = \sum_{k=1}^{\ell-p}\left(\alpha_k - \gamma_k\right)^2 + \sum_{k+\ell-p+1}^{N}\alpha_k^2 \tag{C-13b}$$

For the *GLRT*, M is not available and must be estimated. The estimate is given by equation (C-7), which for the case of interest here ($V = \sigma^2 I$) is equivalently the least-squares estimate or the maximum likelihood estimate. Since the estimate of M must minimized the sum of the squared error in the fit of the mean, M, to be predicted residuals, $\tilde{Z}$, it follows that M must minimize equation (C-13). From the right-hand-side of equation (C-13) the least squares estimate of $\hat{M} = R\hat{\gamma}$ is

$$H_i: \hat{\gamma}_k = \alpha_k; k = 1,\ldots,\ell - p, \tag{C-14a}$$

$$H_j: \hat{\gamma}_k = \alpha_k; k = 1,\ldots,\ell. \tag{C-14b}$$

Consequently, equation (C-13) reduces to

$$\|\tilde{Z} - \hat{M}_j\|^2 = \sum_{k=\ell+1}^{N}a_{k^2}, \tag{C-15a}$$

and

$$\|\tilde{Z} - \hat{M}_i\|^2 = \sum_{k=\ell-p+1}^{N}a_{k^2}. \tag{C-15b}$$

Substituting equation (C-15) into the *GLRT* of equation (C-7) and reducing, yields

$$\ell_m\left(\tilde{Z}\right) = \left(\sum_{k=\ell-p+1}^{\ell}\frac{\alpha^2_k}{\sigma^2}\right) \begin{array}{c}\overline{H}_i \\ > \\ < \\ \overline{H}_i\end{array} \ln\lambda^{-2} \tag{C-16}$$

The predicted residual $\tilde{Z}$ is a Gaussian random vector with mean M_i and covariance $\sigma^2 I$, and probability density function (PDF) given by

$$p(\widetilde{Z}) \Rightarrow N(M_i, \sigma^2 I). \tag{C-17}$$

Using equation (C-13) the PDF of the random vector $\underline{\alpha}$ is

$$p(\underline{\alpha}) = [det|R^{-1}|]^{-1} p(R^{-1}\widetilde{Z}). \tag{C18}$$

Since R is a linear orthogonal transformation ($R^{-1} = R^T$ and $det|R| = 1$), it follows that the random vector $\underline{\alpha}$ is also Gaussian with PDF,

$$p(\underline{\alpha}) \Rightarrow N(\gamma, \sigma^2 I). \tag{C19}$$

Consequently, the *GLRT* test statistic ℓ_m is the sum of p squared zero mean, unit variance, independent Gaussian random variables and is, therefore, a chi-squared random variable with p-degrees (rank of the constraint matrix) of freedom. Recall that this result was obtained under the condition that H_i is true. Therefore, the test statistics, ℓ_m, will exceed the threshold and reject H_i when H_i is true with a probability that is given by the chi-squared distribution with p-degrees of freedom. Similarly, if H_j is true, then at least some of the random variables $\{ \alpha_k : k = \ell - p, \dots \ell \}$ are not zero mean and the test statistic becomes a non-central chi-squared random variable with p-degrees of freedom. Under the conditions of H_j, the test statistic increases and, thereby, increases the probability that the threshold is exceeded and H_i is rejected.

The *GLRT* test statistic can be represented in a form that is more convenient to use. Noting from equation (C-12) and (C-13) that

$$\hat{M}_j = \sum_{k=1}^{\ell} \alpha_k r_k \tag{C-20a}$$

and

$$\hat{M}_i = \sum_{k=1}^{\ell-p} \alpha_k r_k \tag{C-20b}$$

it follows that difference in the estimates of the means under hypothesis j and i is

$$\left[\hat{M}_j - \hat{M}_i \right] = \sum_{k=\ell-p+1}^{\ell} \alpha_k r_k , \tag{C-20c}$$

and the norm of the squared distance of the mean is

$$\left\| \hat{M}_j - \hat{M}_i \right\|^2 = \sum_{k=\ell-p+1}^{\ell} \alpha_k^2 . \tag{C-21}$$

Note, however, that the right-handed side of equation (C-21) is the *GLRT* test statistic; hence

$$\|\tilde{Z} - \hat{M}_j\|^2 - \|\tilde{Z} - \hat{M}_i\|^2 = \|\hat{M}_j - \hat{M}_i\|^2 , \tag{C-22a}$$

or

$$\ell_m(\tilde{Z}) = \frac{\|\hat{M}_j - \hat{M}_i\|^2}{\sigma_n^2} \overset{\overline{H}_i}{\underset{\overline{H}_j}{\gtrless}} \ln\lambda^{-2} . \tag{C-22b}$$

That is, the *GLRT* test statistic reduces the norm of the squared distance between the two mean vectors estimated under the *j*th and *i*th hypothesis. Given that H_j is the full-order model, then ℓ_m is (central) chi-squared with *p*-degrees of freedom when H_i is true and non-central chi-squared with *p*-degrees of freedom when H_j is true.

For specific case of $j = 3$, the tests of interest are

$$T_2: \frac{\|\hat{M}_3 - \hat{M}_2\|^2}{\sigma_n^2} \overset{\overline{H}_2}{\underset{\overline{H}_3}{\gtrless}} \in_2 , \tag{C-23a}$$

$$T_1: \frac{\|\hat{M}_3 - \hat{M}_1\|^2}{\sigma_n^2} \overset{\overline{H}_1}{\underset{\overline{H}_3}{\gtrless}} \in_1 , \tag{C-23b}$$

$$T_0: \frac{\|\hat{M}_3\|^2}{\sigma_n^2} \overset{\overline{H}_0}{\underset{\overline{H}_3}{\gtrless}} \in_0 . \tag{C-23c}$$

The tests T_i evaluate the statistical significance of the presence of each coefficient or feature in the polynomial indicative of anomalies in the predicted residual sequence. For T_2 the test statistic is the norm-squared value in units of noise standard deviation of the difference between the estimates of the mean obtained under the hypothesis that jump and drift features are present ($M_3 \Rightarrow a_0 \neq 0$, $a_1 \neq 0$) and the hypothesis that only drift is present ($M2 \Rightarrow a_0 = 0$, $a_1 \neq 0$). Thus, T_2 is a test of the significance of the non-zero estimate of the jump coefficient a_0. If the test statistic is below threshold, the hypothesis that jump and drift are present is rejected in favor of the hypothesis that only a drift feature is present and no significant jump feature is evident. Conversely, if the test statistic exceeds the threshold, the hypothesis that only a drift is present

(H_2) is rejected in favor of the hypothesis that both jump and drift features (H_3) are evident in the sequence. When H_2 is true, which implies a_0 is zero, then the test statistic is (central) chi-squared with one degree of freedom. Consequently, the threshold, ϵ_2, can be established to give a desired probability of false detection of the jump feature $(a_0 \neq 0)$. When H_3 is true, a jump is present, the jump coefficient a_0 is non-zero, and the test statistic is non-central chi-squared with one-degree of freedom. Although the value of the jump coefficient a_0 is unknown, the parameter of non-centrality could be computed for some value of a_0 and the threshold ϵ_2 set to achieve a specified probability of jump detection at a given amplitude. However, this approach was not used. Notice that rejection of the jump plus drift hypothesis (H_3) does not necessarily imply that the drift hypothesis is accepted. The only conclusion that can be drawn from test T_2 is whether or not a jump is present. Similar comments apply for test T_1 which determines if a drift is present. The last test, T_3, determines if the combined hypothesis of jump plus drift is to be accepted/rejected over the null hypothesis of no anomaly, H_0. This test statistic is chi-squared with two degrees of freedom when H_0 is true. Test T_0 is redundant with tests T_1 and T_2, and is not used.

To complete the hypothesis test, a logic tree is constructed from the following table of possible outcomes.

Table C-1. Multiple Hypotheses Test

T_1 \\ T_2	$(a_0 \neq 0)$ $\overline{H}_2$	$(a_0 = 0)$ $\overline{H}_3$
$(a_1 \neq 0)$ $\overline{H}_1$	$\overline{H}_3$	$\overline{H}_2$
$(a_1 = 0)$ $\overline{H}_3$	$\overline{H}_1$	$\overline{H}_3$

If the result of T_1 indicates the presence of a drift (reject $H_1 \Rightarrow a_1 \neq 0$) and that of T_2 indicates the absence of a jump (reject $H_3 \Rightarrow a_0$), the logical conclusion is that the predicted residuals exhibit a drift-only feature.

C-8

FEATURE AMPLITUDE EVIDENCE REPRESENTATIONS

Evidence pertaining to the amplitude of a feature is embedded in the probability density function of the feature estimates. The feature estimates are obtained from equation (22) of the text and the density function approximated by a Gaussian density; the mean value of the density is taken a the estimated value and the variance obtained from equation (22). Basic probability assignments for the feature amplitude frame of discernment are obtained from operations on the expected value of the feature amplitudes as summarized below. Feature amplitude evidence representations are discussed in greater detail by Ferkinhoff, et al.[14] cited in the text.

Figure D-1 illustrates the feature density function (lower portion) and the feature amplitude membership functions. The first step is to compute the expected for each of the feature amplitudes: strong negative (s^-), moderate negative (m^-), weak negative or positive ($w^{\pm}$, moderate positive (m^+), and strong positive (s^+).
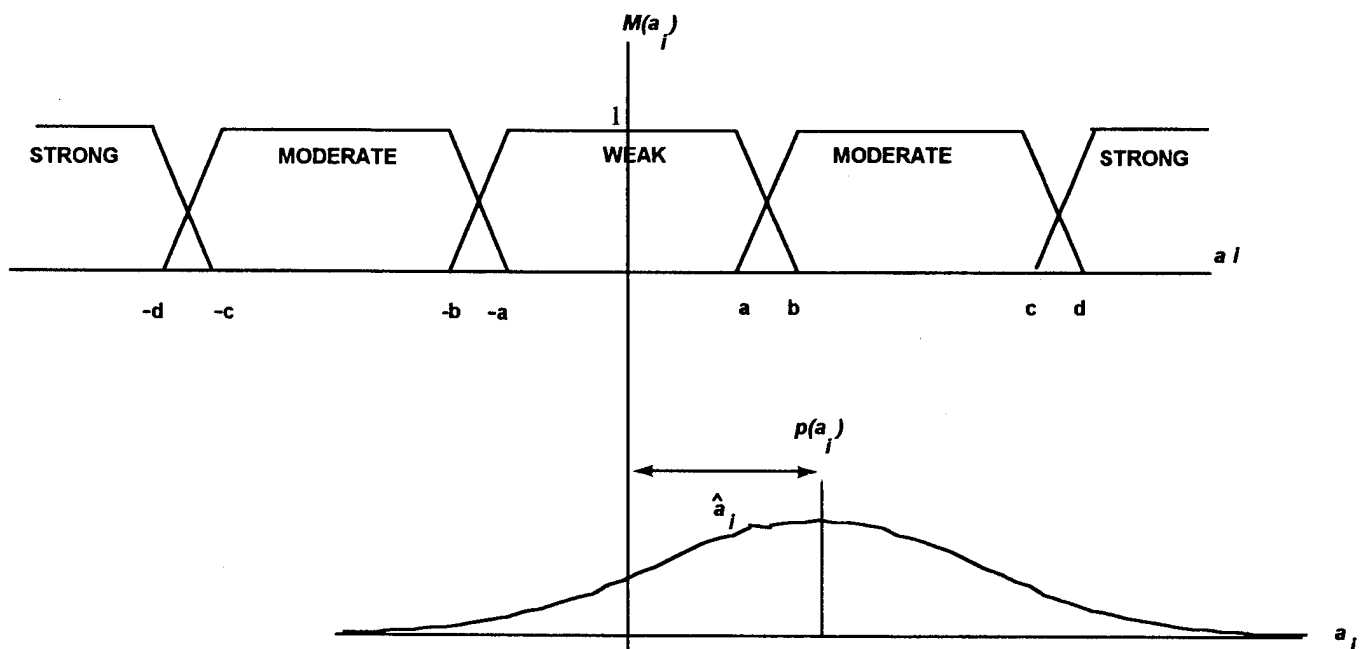

Figure D-1. Feature Amplitude Regions With Polarity Discrimination

The expected values for the amplitudes with polarity are given equations D-1 and D-2.

$$E\{w \pm (a_i)\} = \int_{-\infty}^{+\infty} M_w(a_i)P(a_i)da_i ,$$ (D-1a)

$$E\{m_m + (a_i)\} = \int_{-\infty}^{+\infty} M_{m+}(a_i)P(a_i)da_i ,$$ (D-1b)

$$E\{m_m - (a_i)\} = \int_{-\infty}^{+\infty} M_{m-}(a_i)P(a_i)da_i ,$$ (D-1c)

$$E\{m_s^{+} (a_i)\} = \int_{-\infty}^{+\infty} M_{s+}(a_i)P(a_i)da_i ,$$ (D-1d)

$$E\{m_s^{-} (a_i)\} = \int_{-\infty}^{+\infty} M_{s-}(a_i)P(a_i)da_i .$$ (D-1e)

where the Mj are the individual membership functions for weak, moderate positive, moderate negative, strong positive, and strong negative, respectively.

For unimodal density functions some elements of the power set for the feature amplitude frame of discernment with polarity result in illogical conclusions. For example, the set {strong negative, strong positive} is not a logical possibility. A feature has either a strong positive amplitude or a strong negative amplitude, but not both simultaneously. Similarly, the conditions of {weak, strong positive} cannot exist. Consequently, a reduced frame of discernment is considered for implementation as show in figure D-2.
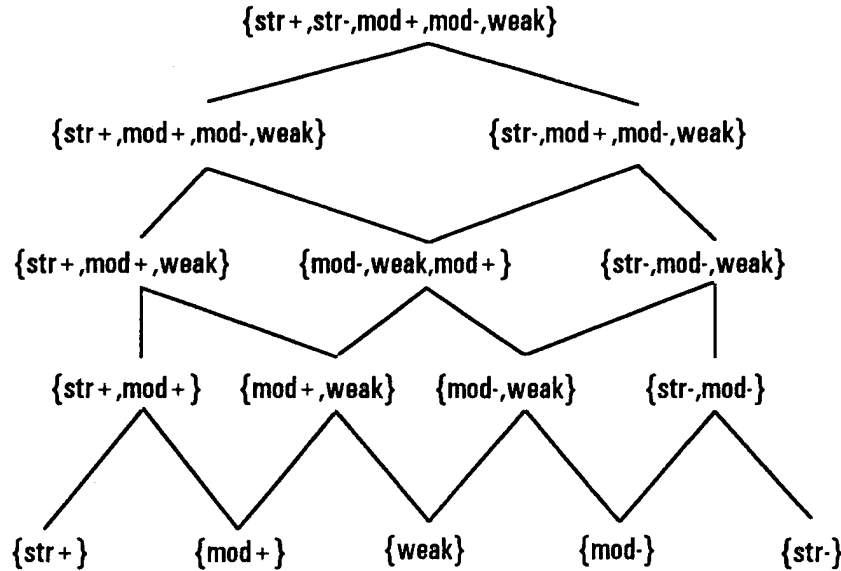

Figure D-2. Revised Power Set for Discriminating Polarity

The basic probability assignments for the elements of the power set of the feature amplitude frame of discernment with polarity is illustrated in figure D-3.
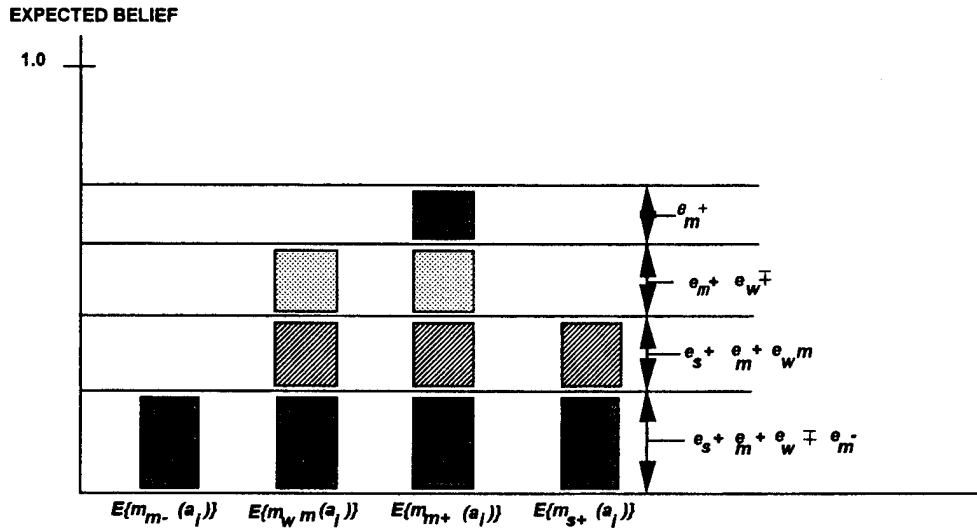

Figure D-3. Belief Distribution When Discriminating Polarity

Here, figure D-3 illustrates the expected values of the feature amplitudes as the vertical bars of a histogram. The BPA is determined from a form of signal excess similar to that for feature existence as explained in the text. The histogram is cut at the various values for the expected values. The BPA is the difference in levels times the number of bars in the histogram that are intersected. The resulting value of BPA is assigned to the union of the sets represented in the cut. Thus, for the example illustrated in figure D-3, the twice the difference between the expected values for moderate positive feature and a weak feature is the BPA given to set {weak, moderate positive}. The equations for the complete assignment of BPA to the reduced frame are given below in equations (D-2)

$$m_s^+ = MAX[e_s^+(a_i) - MAX[e_m^+(a_i),e_m^-(a_i),e_s^-(a_i),e_w^\pm(a_i)] , 0.0] ,$$ (D-2a)

$$m_s^- = MAX[e_s^-(a_i) - MAX[e_m^+(a_i),e_m^-(a_i),e_s^+(a_i),e_w^m(a_i)] , 0.0] ,$$ (D-2b)

$$m_m^+ = MAX[e_m^+(a_i) - MAX[e_m^-(a_i),e_s^+(a_i),e_s^-(a_i),e_w^m(a_i), , 0.0] ,$$ (D-2c)

$$m_m^- = MAX[e_m^-(a_i) - MAX[e_m^+(a_i),e_s^+(a_i),e_s^-(a_i),e_w^m(a_i)] , 0.0] ,$$ (D-2d)

$$m_w^\pm = MAX[e_w^m(a_i) - MAX[e_m^+(a_i),e_m^-(a_i),e_s^+(a_i),e_s^-(a_i)] , 0.0] ,$$ (D-2e)

$$m_{s\,m}^{+\,+} = MAX[2.0\{MIN[e_s^+(a_i),e_m^+(a_i)]-MAX[e_w^m(a_i),e_m^-(a_i),e_s^-(a_i)]\},0.0],$$ (D-2f)

$$m_{m\,w}^{+\,\pm} = MAX[2.0\{MIN[e_m^+(a_i),e_w^\pm(a_i)]-MAX[e_s^+(a_i),e_m^-(a_i),e_s^-(a_i)]\},0.0],$$ (D-2g)

$$m_{m\,w}^{-\,\pm} = MAX[2.0\{MIN[e_m^-(a_i),e_w^\pm(a_i)]-MAX[e_s^+(a_i),e_m^+(a_i),e_s^-(a_i)]\},0.0],$$ (D-2h)

$$m_{s\,m}^{-\,-} = MAX[2.0\{MIN[e_s^-(a),e_m^-(a)]-MAX[e_w^\pm(a),e_m^+(a),e_s^+(a)]\},0.0],$$ (D-2i)

$$m_s^{+}{}_m^{+}{}_w^{\pm} = MAX[3.0\{MIN[e_s^{+}(a),e_m^{+}(a),e_w^{\pm}(a)]-MAX[e_s^{-}(a),e_m^{-}(a)]\},0.0],$$ (D-2j)

$$m_m^{+}{}_w^{\pm}{}_m^{-} = MAX[3.0\{MIN[e_m^{-}(a_i),e_m^{+}(a_i),e_w^{\pm}(a_i)]-MAX[e_s^{-}(a_i),e_s^{+}(a_i)]\},0.0],$$ (D-2k)

$$m_s^{-}{}_m^{-}{}_w^{\pm} = MAX[3.0\{MIN[e_s^{-}(a_i),e_m^{-}(a_i),e_w^{\pm}(a_i)]-MAX[e_s^{+}(a_i),e_m^{+}(a_i)]\},0.0],$$ (D-2l)

$$m_s^{+}{}_m^{+}{}_w^{\pm}{}_m^{-} = MAX[4.0\{MIN[e_s^{+}(a_i),e_m^{+}(a_i),e_w^{\pm}(a_i),e_m^{-}(a_i)]-e_s^{-}(a_i)\}, 0.0],$$ (D-2m)

$$m_s^{-}{}_m^{+}{}_w^{\pm}{}_m^{-} = MAX[4.0\{MIN[e_s^{-}(a_i),e_m^{+}(a_i),e_w^{\pm}(a_i),e_m^{-}(a_i)]-e_s^{+}(a_i)\}, 0.0],$$ (D-2n)

$$m_s^{+}{}_m^{+}{}_w^{\pm}{}_m^{-}{}_s^{-} = 5.0MIN[e_s^{+}(a_i),e_m^{+}(a_i),e_w^{\pm}(a_i),e_m^{-}(a_i),e_s^{-}(a_i)] .$$ (D-2o)

Since the amplitude frame does not distinguish between positive and negative features the final BPA assignments are without regard to polarity and are given by the compatibility relation shown in table D-1.

Table D-1. Compatibility Relation Between Feature Amplitudes
With and Without Polarity

Feature Amplitude Frame With Polarity	Feature Amplitude Frame Without Polarity
$\{jw,-jm,+jm,-js,+js\}$	$\{jw,jm,js\}$
$\{jw,-jm,+jm,-js\}$	$\{jw,jm,js\}$
$\{jw,-jm,+jm,,+js\}$	$\{jw,jm,js\}$
$\{jw,-jm,-js\}$	$\{jw,jm,js\}$
$\{jw,+jm,+js\}$	$\{jw,jm,js\}$
$\{jw,-jm,+jm,\}$	$\{jw,jm\}$
$\{jw,-jm\}$	$\{jw,jm\}$
$\{jw,+jm\}$	$\{jw,jm\}$
$\{-jm,-js\}$	$\{jm,js\}$
$\{+jm,+js\}$	$\{jm,js\}$
$\{+js\}$	$\{js\}$
$\{-js\}$	$\{js\}$
$\{-jm\}$	$\{jm\}$
$\{+jm\}$	$\{jm\}$
$\{jw\}$	$\{jw\}$

Note that for single mode probability density functions, like the Gaussian, logically inconsistent assignments of BPA to sets like $\{-js,+js\}$ or $\{jw,-js\}$ cannot occur with this technique. Consequently, non-zero BPA for both sets $\{-js\}$ and $\{+js\}$ also cannot occur simultaneously. However, a non-zero BPA assignment for the sets $\{-js\}$ and $\{-js,-jm,jw,+jm,+js\}$ is possible and does not create a logically inconsistent result.

A total BPA of unity in the frame of discernment is obtained only if special symmetry relations are maintained with the functions that define the regions of feature amplitudes in figure 7b. The boundaries for the feature amplitude regions were obtained from a perturbation analysis. The perturbation analysis was performed numerically with a monte carlo simulation and the results presented in the form of scatter plots of event occurrence in a jump-drift feature plane. The scatter plots are discussed in appendix E. The boundaries were established by visual inspection; no attempt was made to optimize the boundary locations.

APPENDIX E
FEATURE SCATTER PLOTS

Scatter plots were used to determine where the thresholds for the features of strong, moderate, and weak should be set to discriminate among the models for target maneuver, change in propagation path, change in base frequency, and unknown model changes. Additionally, the scatter plots were also used to determine the compatibility relationships. The scatter plots are shown in figures E-1 through E-4 for the conical angle, sphere bearing angle, sphere depression/elevation angle, and frequency, respectively. Only one frequency plot for both sensors is presented because the results are practically identical.

In all the plots, the jump amplitude is plotted along the horizontal axis; the drift amplitude is plotted along the vertical axis; and the location of the thresholds have been indicated with solid lines. The individual values of the jump and drift coefficients are indicated on the plot using three different symbols, a solid square for contact maneuver (CM), a diamond for change in propagation path (PP), and a circle for a change in the base frequency (BF). The plots were obtained from a monte carlo simulation involving 42,000 trials. The measurement data for each trail was generated from a randomized single leg target-observer geometry with 23 measurement sets at 20-second intervals. The randomized geometries were selected using a uniform distribution of relative contact course values of five samples between 0° and 350°; relative contact speed of five uniformly selected samples between 2 and 15 yd/sec; frequency five samples uniformly selected between 200 Hz and 2 kHz; depth between 100 yd and 10 kyd; and a nominal sound speed of 1400 yd/sec. The model anomalies were target course changes of ±170° in six divisions; speed change of ±10 yd/sec in six divisions; depth changes of ±10 kyd in 15 divisions; and frequency changes of 0.5 percent to 1.1 percent in 15 divisions.

For each of the three types of model anomalies, CM, PP, BF, 14,000 monte carlo trials were obtained. For each monte carlo trial, a random geometry was generated and the noise-free predicted residuals were computed. The linear regression fit to the predicted residuals was obtained yielding the jump and drift coefficients. The values of the jump and drift coefficients were quantized and placed in bins and accumulated, by bin, across the trials. All the bins with non-zero entries are shown in the plots E-1 through E-4. Note that in these scatter plots, bins with one entry are represented equally with bins that have thousands of entries. The thresholds are set empirically using visual inspection to obtain a meaningful separation among the classes of models. No attempt was made to optimize this the selection of the boundaries. The thresholds obtained by this procedure are given in tables within the plots E-1 through E4 for the azimuth bearing, DE angle, conical angle, and frequency, respectively.
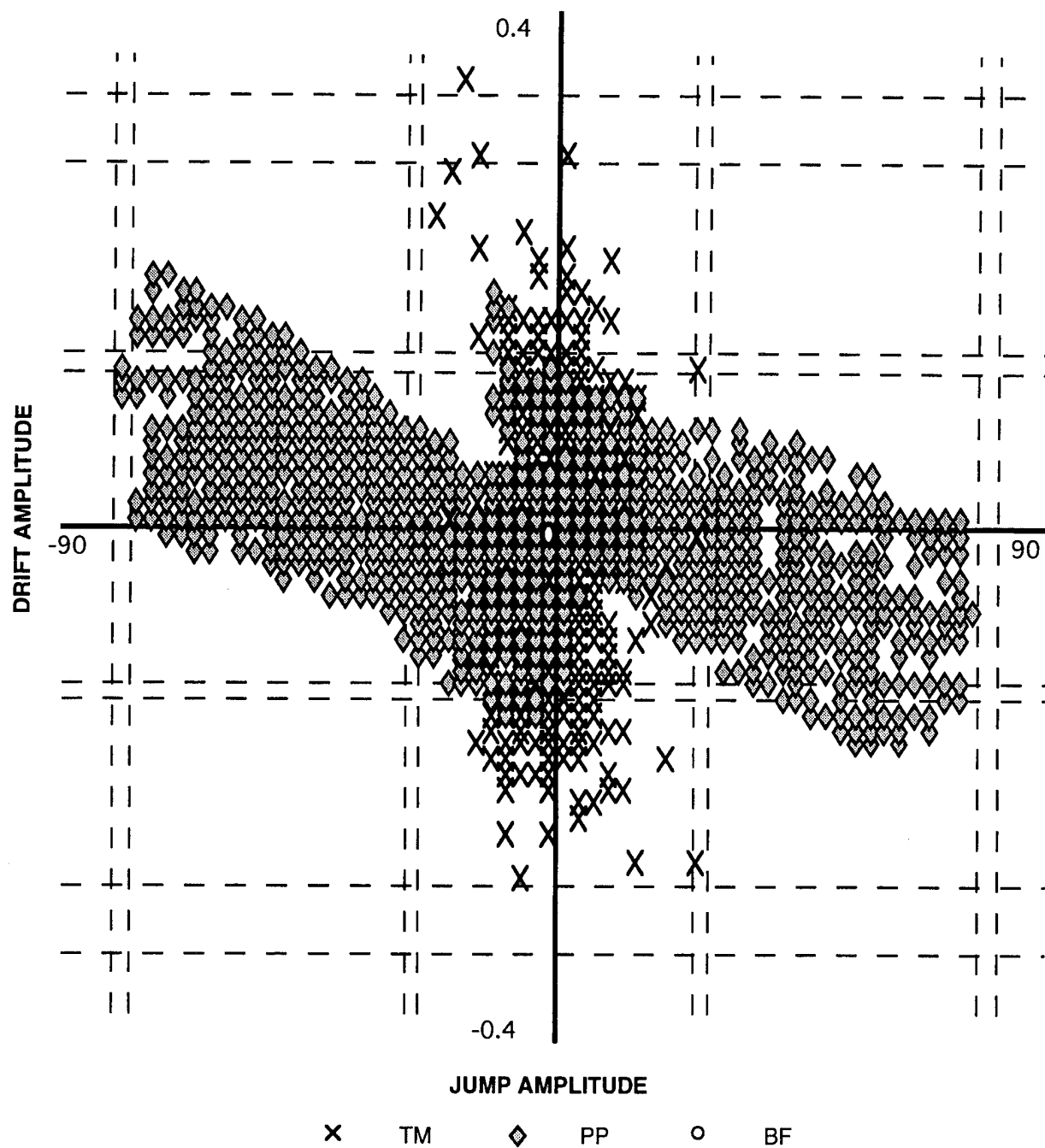
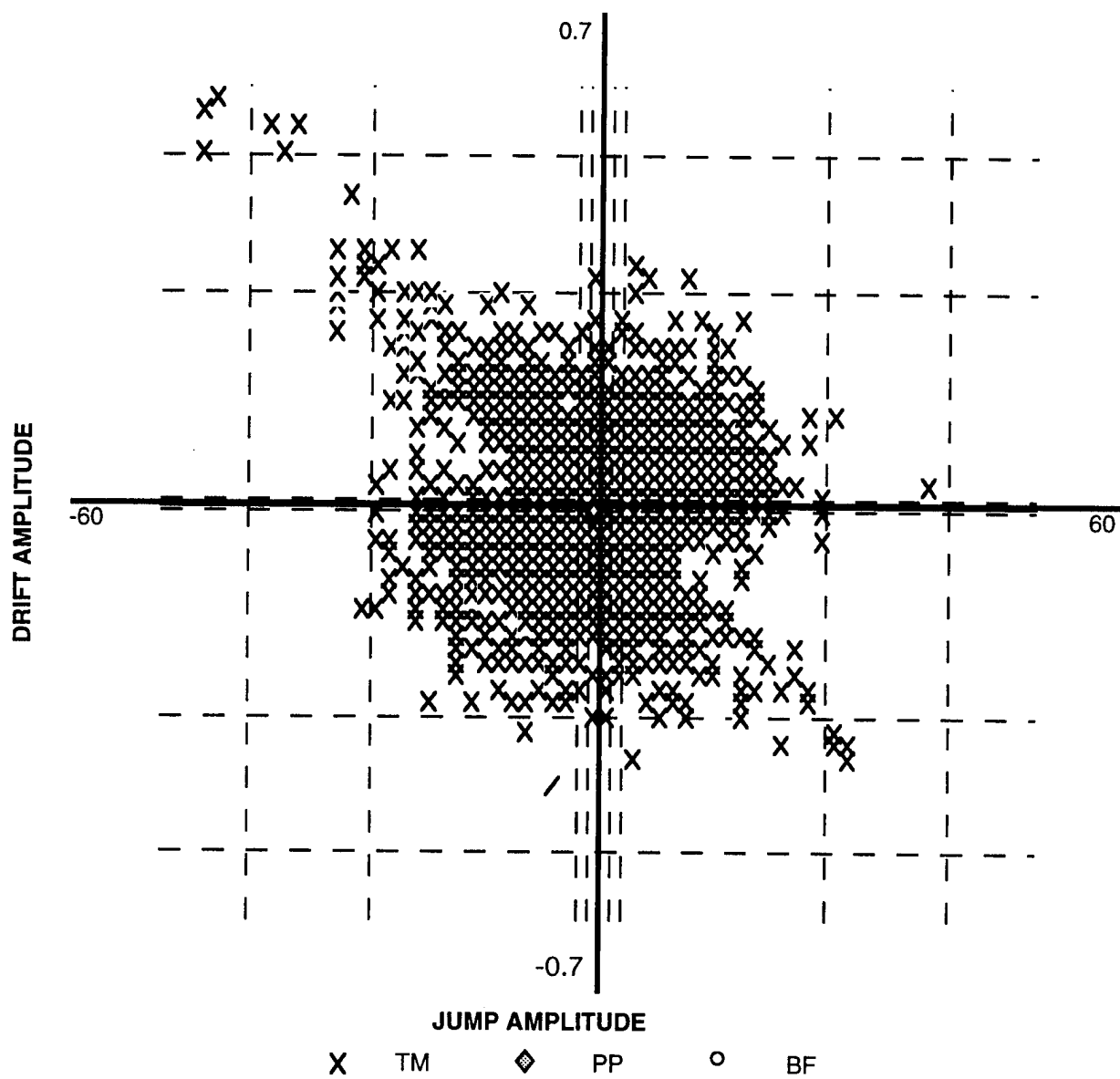*Figure E-1. Scatter Plot for Conical Angle*
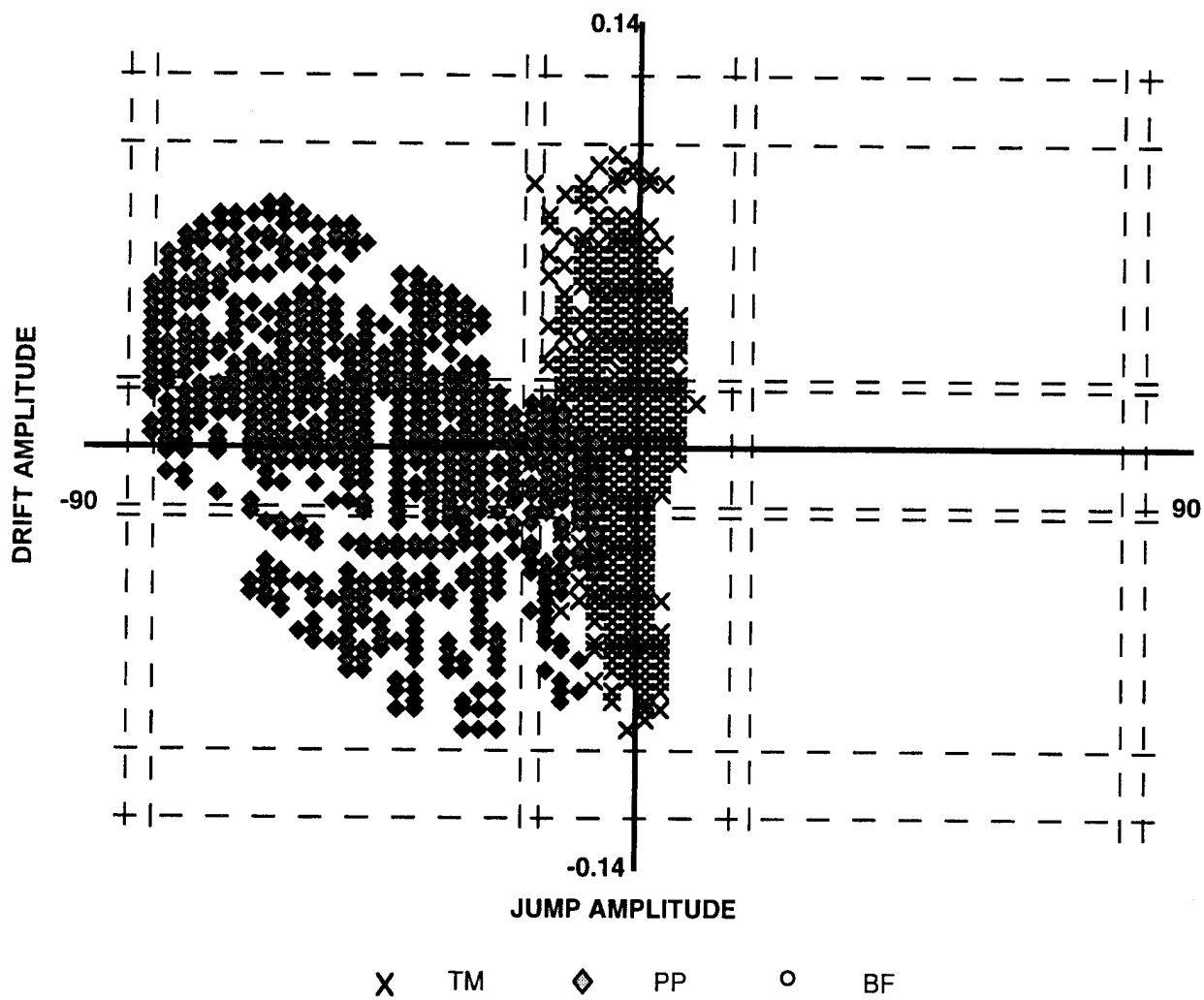
Figure E-2. Scatter Plot for Sphere Bearing Angle

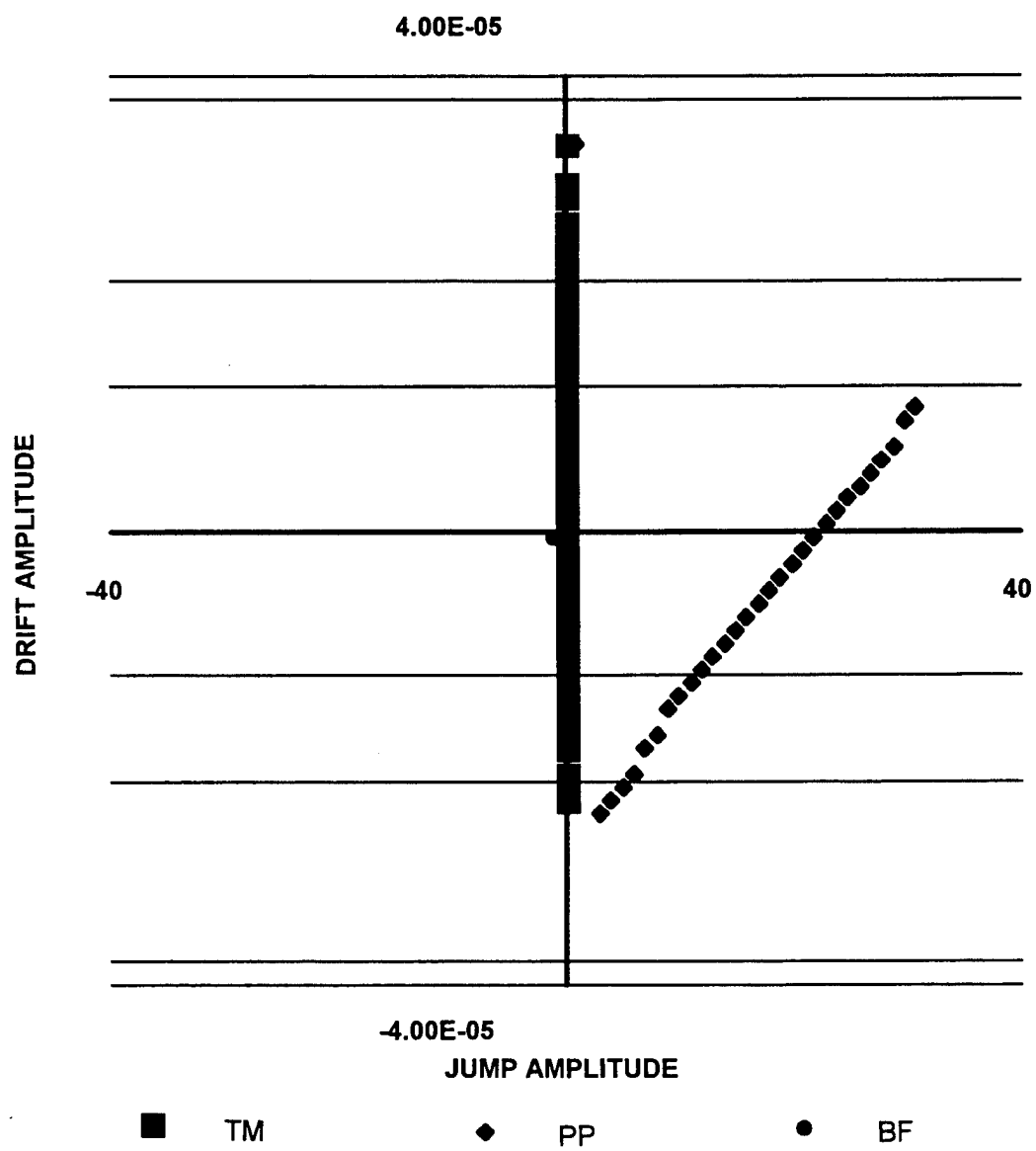*Figure E-3. Scatter Plot for Sphere Depression/Elevation Angle*

Figure E-4. Scatter Plot for Frequency

REFERENCES

1. Y. T. Chan and J. J. Towers, "Sequential Localization of a Radiating Source by Doppler-Shifted Frequency Measurements", *IEEE Transactions on AES*, vol. 28, no. 4, October 1992, pp. 1084-1090.

2. Y. T. Chan and S. W. Rudnicki, "Bearings-Only and Doppler-Shifted Tracking Using Instrumental Variables," *IEEE Transactions on AES*, vol. 28, no. 4, October 1992, pp. 1076-1083.

3. M. Gauish and A. J. Weiss, "Performance Analysis of Bearings-Only Target Localization Algorithms," *IEEE Transactions on AES*, vol. 28, no. 4, July 1992, pp. 817-828.

4. A. Holtsberg, "A Statistical Analysis of Bearings-Only Tracking," Doctoral Dissertation, Lund Institute of Technology, Lund, Sweden, 1992.

5. J. G. Baylog, A. A. Magliaro, S. M. Zile, and K. F. Gong, "Underwater Tracking in the Presence of Modeling Uncertainty," *Proceedings of the Twenty-First Asilomar Conference on Signals Systems and Computers*, November 1987.

6. D. Kolb and F. Holister, "Bearings-Only Contact Parameter Estimation," *Proceedings of the First Asilomar Conference on Signals Systems and Computers*, November 1967.

7. K. F. Gong, S. E. Hammel, S. C. Nardone, and A. G. Lindgren, "Three Dimensional Contact Parameter Estimation," *Proceedings of the Sixteenth Asilomar Conference on Signals Systems and Computers*, November 1982.

8. K. F. Gong, J. G. Baylog, and A. A. Magliaro, "A Decision Directed Approach to Solution Integration for Tracking in an Underwater Environment," *Proceedings of the Eighteenth Asilomar Conference on Signals Systems and Computers*, November 1984.

9. D. J. Ferkinhoff, J, G, Baylog, K. F. Gong, and S. C. Nardone, "Feature Extraction and Interpretation for Dynamic System Model Resolution," *Proceedings of the Twenty-Fourth Asilomar Conference on Signals Systems and Computers*, November 1990.

10. D. J. Ferkinhoff, K. F. Gong, K. D. Keay, J. F. MacDonald, and S. C. Nardone, "Evidence Generation and Representation for Model Uncertainty Management in Nonlinear State Estimation Systems," *Proceedings of the Twenty-Fifth Asilomar Conference on Signals Systems and Computers*, November 1991.

11. A. P. Dempster, "A Generalization of Bayesian Inference," *Journal of the Royal Statistical Society*, Series B, 30, pp 205-247, 1968.

12. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.

13. D. J. Ferkinhoff, J. G. Baylog, K. F. Gong. and S. C. Nardone, "Feature Detection for Model Assessment in State Estimation," Technical Report 7064, Naval Underwater Systems Center, Newport, RI, 1991 (UNCLASSIFIED).

14. D. J. Ferkinhoff et. al., "Functions for Mapping Statistical Multiple Hypothesis Test Results to Belief for Evidential Reasoning," Technical Report 10,286, Naval Undersea Warfare Center, Newport, RI, June 1993 (UNCLASSIFIED).

15. J. D. Lowrance, T. D. Garvey, and T. M. Strat, "A Framework for Evidential Reasoning Systems," *Proceedings AAAI-86 Fifth National Conference on Artificial Intelligence*, 11-15 August 1986, Philadelphia, PA, vol. I.

16. M. L. Graham, *A Hierarchical Approach to Estimation, With Application to Target Motion Analysis*, PhD Dissertation, University of Rhode Island, 1992.

17. M. L. Graham, K. D. Keay, J. F. MacDonald, and K. F. Gong, "Data Fusion and Tracking with Uncertain Modeling and Data Origin," 1992 BRGC^2 *Symposium on Command and Control Research*, Montery CA. June 1992.

18. M.L. Graham and F. O'Brien, "Rank Analysis Module for Target Motion Analysis", Patent Application, Navy Case 72911, 20 June 1990.

19. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley & Sons, New York, 1968.

20. P. Kersten et. al., (Technical Paper in Preparation) 1993.

21. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

22. A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics, 5th Edition Vol. 2.*, Oxford University Press, NY, 1991.

23. P. G. Hoel, S. C. Port, and C. J. Stone, *Introduction to Statistical Theory*, Houghton, Mifflin Company, Boston, MA., 1971.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Office of Naval Research (J. Fein -- ONR-333)	1
Naval Sea Systems Command (ASTO-E)	1
Defense Technical Information Center	12
Center for Naval Analyses	1